

GenCore version 5.1.6  
Copyright (c) 1993 - 2004 Compugen Ltd.

OM protein - protein search, using sw model

Run on: July 20, 2004, 11:10:18 ; Search time 54 Seconds  
(without alignments)  
1198.211 Million cell updates/sec

Title: US-09-942-052a-728  
1198

Perfect score: 1 MAAPLRHRRCATPPRGDF.....LKSIKILSEVTDPQSKPEN 229

Sequence: BLOSUM62  
Gapop 10.0 , Gapext 0.5

Scoring table: 1586107 seqs, 282547505 residues

Searched: Total number of hits satisfying chosen parameters: 1586107

Minimum DB seq length: 0  
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%  
Maximum Match 100%  
Listing first 45 summaries

Database : A\_Geneseq\_29Jan04:\*

1: geneseqp1980s:\*\n2: geneseqp1990s:\*\n3: geneseqp2000s:\*\n4: geneseqp2001s:\*\n5: geneseqp2002s:\*\n6: geneseqp2003as:\*\n7: geneseqp2003bs:\*\n8: geneseqp2004s:\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

## SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	1198	100.0	229	4	AAW40223 Human pol
2	1198	100.0	229	5	ABG34856 Human can
3	1198	100.0	231	4	AAW42009 Human pol
4	138.5	11.6	164	5	ABG34855 Human CDN
5	113	9.4	233	4	AAW93668 Human pol
6	113	9.4	233	4	AAW93290 Human pro
7	92	7.7	635	5	ABBS4167 Lactococc
8	90	7.5	269	4	AAO08972 Human pol
9	89	7.4	623	4	AAW93182 Human pro
10	89	7.4	623	5	ABBS7223 Novel hum
11	89	7.4	652	4	ABBS9168 Human pro
12	89	7.4	664	4	AAW83843 Human ac1
13	89	7.4	664	4	AAW83843 Human ac1
14	87.5	7.3	1245	7	ADG32818 Human nov
15	86.5	7.2	669	6	ABU23878 Protein e
16	84.5	7.1	485	5	ABPS1336 Human MDD
17	84	7.0	496	7	ADG31196 Human nov
18	84	7.0	496	7	ADG31196 Human nov
19	83.5	7.0	479	6	ABM36197 Propionib
20	83.5	7.0	479	6	ABM36197 Propionib
21	83	6.9	598	2	AAW59461 Microbial
22	83	6.9	598	2	AAW59463 Microbial
23	83	6.9	598	2	AAW59457 Microbial
24	83	6.9	598	2	AAW59458 Microbial
25	83	6.9	598	2	AAW59459 Microbial

26	83	6.9	598	2	AAW59462 Microbial
27	83	6.9	598	2	AAW59460 Microbial
28	83	6.9	598	2	AAW59464 Microbial
29	83	6.9	599	2	AAW59496 Microbial
30	83	6.9	606	1	AAW90110 Polypepti
31	83	6.9	607	2	AAW59456 Microbial
32	83	6.9	607	2	AAW59465 Microbial
33	83	6.9	663	6	ABU23749 Protein e
34	82.5	6.9	325	3	AAW90242 Human cyc
35	81.5	6.8	456	1	AAW91895 Protein s
36	81.5	6.8	457	2	AAW42430 Escherich
37	81.5	6.8	701	4	ABG01389 Novel hum
38	80.5	6.7	282	4	ABG20699 Novel hum
39	80.5	6.7	766	5	ABW93270 Novel hum
40	80.5	6.7	766	5	ABW93271 Herbicida
41	80.5	6.7	1711	2	AAW70506 Osteotest
42	80.5	6.7	1711	2	AAW70507 Mutant os
43	80.5	6.7	1711	5	ABW52349 Protein r
44	80	6.7	305	5	AAE21310 Mouse Mrg
45	79.5	6.6	212	3	AAW53428 Human col

## ALIGNMENTS

RESULT 1  
ID AAW40223 standard; protein; 229 AA.  
AC AAW40223;  
DT 22-OCT-2001 (first entry)  
DE Human polypeptide SEQ ID NO 3368.  
XX Human; nocrotic; immunosuppressant; cytosolic; gene therapy; cancer;  
XX peripheral nervous system; neuropathy; central nervous system; CNS;  
XX Alzheimer's; Parkinson's disease; Huntington's disease; haemostatic;  
XX amyotrophic lateral sclerosis; Shy-Drager Syndrome; chemotactic;  
XX chemokinetic; thrombolytic; drug screening; arthritis; inflammation;  
XX leukaemia.  
OS Homo sapiens.  
XX  
XX WO200153312-A1.  
XX  
XX 26-JUL-2001.  
XX  
XX 26-DEC-2000; 2000MO-US034263.  
XX  
XX 23-DEC-1999; 99US-00471275.  
XX 21-JAN-2000; 2000US-00488725.  
XX 25-APR-2000; 2000US-00552317.  
XX 20-JUN-2000; 2000US-00598042.  
XX 19-JUL-2000; 2000US-00620312.  
XX 03-AUG-2000; 2000US-00653450.  
XX 14-SEP-2000; 2000US-00662191.  
XX 19-OCT-2000; 2000US-00693036.  
XX 29-NOV-2000; 2000US-00727344.  
XX  
XX (HYSE-) HYSEQ INC.  
XX  
XX Tang YT, Liu C, Asundi V, Chen R, Ma Y, Qian XB, Ren F, Wang D,  
XX Wang J, Wang Z, Wehrman T, Xu C, Xue AJ, Yang Y, Zhang J, Zhao QA,  
XX Zhou P, Goodrich R, Drmanac R,  
XX WPI, 2001-442253/47.  
XX N-PSDB; AAI59379.  
XX  
XX Novel nucleic acids and polypeptides, useful for treating disorders such  
XX as central nervous system injuries.  
XX  
XX Example 5; SEQ ID NO 3368; 10078pp; English.

XX CC The invention relates to human nucleic acids (AA15798-AA161369) and the  
CC encoded polypeptides (AA038642-AA042213) with nootropic,  
CC immunosuppressant and cytostatic activity. The polynucleotides are useful  
CC in gene therapy. A composition containing a polypeptide or polynucleotide  
CC of the invention may be used to treat diseases of the peripheral nervous  
CC system, such as peripheral nervous injuries, peripheral neuropathy and  
CC localised neuropathies and central nervous system diseases, such as  
CC Alzheimer's, Parkinson's disease, Huntington's disease, amyotrophic  
CC lateral sclerosis, and Shy-Drager Syndrome. Other uses include the  
CC utilisation of the activities such as: Immune system suppression,  
CC Activin/inhibin activity, chemotactic/chemokinetic activity, haemostatic  
CC and thrombolytic activity, cancer diagnosis and therapy, drug screening,  
CC assays for receptor activity, arthritis and inflammation, leukaemias and  
CC C.N.S disorders. Note: The sequence data for this patent did not form  
CC part of the printed specification  
XX  
SQ Sequence 229 AA;  
Query Match 100.0%; Score 1198; DB 4; Length 229;  
Best Local Similarity 100.0%; Pred. No. 5, 2e-127;  
Matches 229; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
QY 1 MAAGPLRHSRCATPPRGDFCGGTERAIDQASFTTSMEMDTQVVGKSSPLGAGLGAEEP 60  
DB 1 MAAGPLRHSRCATPPRGDFCGGTERAIDQASFTTSMEMDTQVVGKSSPLGAGLGAEEP 60  
QY 61 AAGPOLPSMLOPERCAVFOCAQCHAVLADSVHLAMDLSRSLGAVVSRVTNNVLEAPFL 120  
DB 61 AAGPOLPSMLOPERCAVFOCAQCHAVLADSVHLAMDLSRSLGAVVSRVTNNVLEAPFL 120  
QY 121 VGIEGSLKGSTYNNLLFCGSCGIPVGPHLYSTHAALALRGHFCSSDDKWCYLLTKTAIV 180  
DB 121 VGIEGSLKGSTYNNLLFCGSCGIPVGPHLYSTHAALALRGHFCSSDDKWCYLLTKTAIV 180  
QY 181 NASEMDIONVPLSEKIAELKEKIVLTHNRILKSLMKLISEVTPDQSKPEN 229  
DB 181 NASEMDIONVPLSEKIAELKEKIVLTHNRILKSLMKLISEVTPDQSKPEN 229  
RESULT 2  
ABG34856 standard; protein; 229 AA.  
XX  
AC ABG34856;  
XX  
DT 15-JUL-2002 (first entry)  
XX  
DE Human cancer related protein encoded by cDNA 85PIB3.  
XX  
KW Human; cytostatic; 85PIB3; cancer; immunogen; chromosome 15q14.  
XX  
OS Homo sapiens.  
XX  
PN WO200218578-A2.  
XX  
PD 07-MAR-2002.  
XX  
PF 28-AUG-2001; 2001WO-US026838.  
XX  
PR 28-AUG-2000; 2000US-0228432P.  
XX  
PA (AGEN-) AGENSYS INC.  
XX  
PI Raitano AB, Faris M, Hubert RS, Afar D, Ge W, Chalilta-Eid P,  
PI Jakobovits A;  
XX  
DR WPI; 2002-382963/41.  
DR N-PSDB; ABK70506.  
XX  
PT Composition for modulating the status of 85PIB3 protein or a molecule  
PT comprising a substance e.g. antibody specific to, nucleic acid encoding,  
PT or ribozyme of 85PIB3.

XX CC Claim 34; Fig 2; 201pp; English.  
PS  
XX  
XX CC The invention relates to a composition comprising a substance that  
CC modulate the status of 85PIB3, where the status of a cell expresses  
CC 85PIB3 gene product is modulated. Also included are a composition  
CC comprising a peptide region of 5 amino acids of the 85PIB3 protein, in  
CC any whole number increment up to 229 that includes an aa position  
CC selected from an aa position having a value greater than 0.5 in the  
CC Hydrophobicity profile, an aa position having a value less than 0.5 in  
CC the hydrophobicity profile, an aa position having a value greater than  
CC 0.5 in the percent accessible residue profile, an aa position having a  
CC value greater than 0.5 in the average flexibility profile, or an aa  
CC position having a value greater than 0.5 in the beta-turn profile; a  
CC polynucleotide that encodes analogue peptide of 8, 9, 10 or 11 contiguous  
CC residues of the 85PIB3 protein; a recombinant protein comprising the  
CC antigen-binding region of a monoclonal antibody; a non-human transgenic  
CC animal that produces an antibody that binds to the 85PIB3 protein; a  
CC hybridoma that produces antibody specific to the protein; a single chain  
CC monoclonal antibody (MAb) that comprises the variable domains of the  
CC heavy and monoclonal antibodies specific to the protein; a vector  
CC comprising a polynucleotide that encodes the MAb; inhibiting growth of  
CC cancer cells or treating a patient who bears cancer cells that expresses  
CC the protein, by administering the protein, antibody, polynucleotide  
CC encoding the protein, antisense polynucleotide to the polynucleotide,  
CC ribozyme that cleaves the polynucleotide and T cells that specifically  
CC recognize the protein, and generating a mammalian immune response  
CC directed to the protein exposing cells of the mammal's immune system to  
CC an immunogenic portion of the protein or polynucleotide. The composition,  
CC which comprises an antibody specific to the protein, is useful for  
CC delivering a cytotoxic agent to a cell that expresses the protein by  
CC providing a cytotoxic agent conjugated to antibody and exposing the cell  
CC to the antibody-agent conjugate. The methods are useful for inhibiting  
CC growth of cancer cells or treating a patient who bears cancer cells that  
CC expresses the protein, for generating a mammalian immune response  
CC directed to the protein, for detecting the presence of the protein or  
CC polynucleotide in a biological sample in a patient who has or who is  
CC suspected of having cancer and for monitoring 85PIB3 in a biological  
CC sample from a patient who has or who is suspected of having cancer. The  
CC gene for 85PIB3 is located on human chromosome 15q14. The present  
CC sequence is the 85PIB3 protein  
XX  
SQ Sequence 229 AA;  
Query Match 100.0%; Score 1198; DB 5; Length 229;  
Best Local Similarity 100.0%; Pred. No. 5, 2e-127;  
Matches 229; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
QY 1 MAAGPLRHSRCATPPRGDFCGGTERAIDQASFTTSMEMDTQVVGKSSPLGAGLGAEEP 60  
DB 1 MAAGPLRHSRCATPPRGDFCGGTERAIDQASFTTSMEMDTQVVGKSSPLGAGLGAEEP 60  
QY 61 AAGPOLPSMLOPERCAVFOCAQCHAVLADSVHLAMDLSRSLGAVVSRVTNNVLEAPFL 120  
DB 61 AAGPOLPSMLOPERCAVFOCAQCHAVLADSVHLAMDLSRSLGAVVSRVTNNVLEAPFL 120  
QY 121 VGIEGSLKGSTYNNLLFCGSCGIPVGPHLYSTHAALALRGHFCSSDDKWCYLLTKTAIV 180  
DB 121 VGIEGSLKGSTYNNLLFCGSCGIPVGPHLYSTHAALALRGHFCSSDDKWCYLLTKTAIV 180  
QY 181 NASEMDIONVPLSEKIAELKEKIVLTHNRILKSLMKLISEVTPDQSKPEN 229  
DB 181 NASEMDIONVPLSEKIAELKEKIVLTHNRILKSLMKLISEVTPDQSKPEN 229  
RESULT 3  
AAM42009 standard; protein; 231 AA.  
XX  
ID AAM42009  
XX  
AC AAM42009;  
XX  
DT 22-OCT-2001 (first entry)  
XX

DE Human polypeptide SEQ ID NO 6940.

XX Human; noctropic; immunosuppressant; cyrostatic; gene therapy; cancer;  
KM peripheral nervous system; neuropathy; central nervous system; CNS;  
KM Alzheimer's; Parkinson's disease; Huntington's disease; haemostatic;  
KM amyotrophic lateral sclerosis; Shy-Drager Syndrome; chemotactic;  
KM chemokinetic; thrombolytic; drug screening; arthritis; inflammation;  
KM leukaemia.

XX Homo sapiens.

OS WO200153312-A1.

XX 26-JUL-2001.

XX 26-DEC-2000; 2000WO-US034263.

XX 23-DEC-1999; 99US-00471275.

PR 21-JAN-2000; 2000US-00488725.

PR 25-APR-2000; 2000US-00552317.

PR 20-JUN-2000; 2000US-00598042.

PR 19-JUL-2000; 2000US-00620312.

PR 03-AUG-2000; 2000US-00634450.

PR 14-SEP-2000; 2000US-00662191.

PR 19-OCT-2000; 2000US-00693036.

PR 29-NOV-2000; 2000US-00727344.

XX (HYSE-) HYSEQ INC.

PA Tang YF, Liu C, Asundi V, Chen R, Ma Y, Qian XB, Ren F, Wang D,  
PI Wang J, Wang Z, Wehrman T, Xu C, Xue AJ, Yang Y, Zhang J, Zhao QA,  
PI Zhou P, Goodrich R, Drmanac R;  
XX WPI: 2001-442253/47.  
DR N-PDB; AAI61165.

XX Novel nucleic acids and polypeptides, useful for treating disorders such  
PT as central nervous system injuries.

XX Example 2; SEQ ID NO 6940; 10078pp; English.

XX The invention relates to human nucleic acids (AAI57798-AAI61369) and the  
CC encoded polypeptides (AAM8642-AAM4213) with noctropic,  
CC immunosuppressant and cyrostatic activity. The polynucleotides are useful  
CC in gene therapy. A composition containing a polypeptide or polynucleotide  
CC of the invention may be used to treat diseases of the peripheral nervous  
CC system, such as peripheral nervous injuries, peripheral neuropathy and  
CC localised neuropathies and central nervous system diseases, such as  
CC Alzheimer's, Parkinson's disease, Huntington's disease, amyotrophic  
CC lateral sclerosis, and Shy-Drager Syndrome. Other uses include the  
CC utilisation of the activities such as: Immune system suppression,  
CC Activin/inhibin activity, chemotactic/chemokinetic activity, haemostatic  
CC and thrombolytic activity, cancer diagnosis and therapy, drug screening,  
CC assays for receptor activity, arthritis and inflammation, leukaemia and  
CC C.N.S disorders. Note: The sequence data for this patent did not form  
CC part of the printed specification

XX Sequence 231 AA:

SQ

Query Match 100.0%; Score 1198; DB 4; Length 231;  
Best Local Similarity 100.0%; Pred. No. 5.3e-127;  
Matches 229; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 MAAPLRRHRCATPPRGDFCGGTERAIDQASFTTSMWMDQVVGSSPLGAGAEPP 60  
DB 3 MAAPLRRHRCATPPRGDFCGGTERAIDQASFTTSMWMDQVVGSSPLGAGAEPP 62  
QY 61 AAGQPLPSMLPERRCAVFOCAQCAVAVLADSVHLAMDLSRSI GA VFSVTNNVLEAPFL 120  
DB 63 AAGQPLPSMLPERRCAVFOCAQCAVAVLADSVHLAMDLSRSI GA VFSVTNNVLEAPFL 122  
QY 121 VGIEGSLKSTYNNLFCGSCGIPVGFHLVSTHAALALRGHFCLSDDKMCYLLKTKAIV 180

DB 123 VGIEGSLKSTYNNLFCGSCGIPVGFHLVSTHAALALRGHFCLSDDKMCYLLKTKAIV 182

QY 181 NASEMDIQNVPLSEKIAELKEKIVLTNNRLSKLMLKLTSEVTPDDSKPEN 229  
DB 183 NASEMDIQNVPLSEKIAELKEKIVLTNNRLSKLMLKLTSEVTPDDSKPEN 231

RESULT 4  
ABG34855  
ID ABG34855 standard; protein; 164 AA.

XX ABG34855;

AC 15-JUL-2002 (first entry)

XX Human CDNA 85PIB3 splice variant, open reading frame #3.

DE Human; cyrostatic; 85PIB3; cancer; immunogen; chromosome 15q14.

XX Homo sapiens.

OS WO200218578-A2.

XX 07-MAR-2002.

XX 28-AUG-2001; 2001WO-US026838.

XX 28-AUG-2000; 2000US-0228432P.

XX (AGEN-) AGENSYS INC.

PA Raitano AB, Paris M, Hubert RS, Afar D, Ge W, Challita-Eid P;  
PI Jakobovits A;  
XX WPI: 2002-382963/41.  
DR N-PDB; ABK70504.

XX Composition for modulating the status of 85PIB3 protein or a molecule  
PT comprising a substance e.g. antibody specific to, nucleic acid encoding,  
PT or ribozyme of 85PIB3.

XX Example 38; Page 124; 201pp; English.

XX The invention relates to a composition comprising a substance that  
CC modulate the status of 85PIB3, where the status of a cell expresses  
CC 85PIB3 gene product is modulated. Also included are a composition  
CC comprising a peptide region of 5 amino acids of the 85PIB3 protein, in  
CC any whole number increment up to 229 that includes an aa position  
CC selected from an aa position having a value greater than 0.5 in the  
CC hydrophobicity profile, an aa position having a value less than 0.5 in  
CC the hydrophobicity profile, an aa position having a value greater than  
CC 0.5 in the percent accessible residue profile, an aa position having a  
CC value greater than 0.5 in the average flexibility profile, or an aa  
CC position having a value greater than 0.5 in the beta-turn profile; a  
CC polynucleotide that encodes analogue peptide of 8, 9, 10 or 11 contiguous  
CC residues of the 85PIB3 protein; a recombinant protein comprising the  
CC antigen-binding region of an antibody that binds to the 85PIB3 protein; a  
CC animal that produces an antibody specific to the protein; a single chain  
CC hybridoma that produces an antibody specific to the protein; a vector  
CC monoclonal antibody (Mab) that comprises the variable domains of the  
CC heavy and monoclonal antibodies specific to the protein; a vector  
CC comprising a polynucleotide that encodes the Mab; inhibiting growth of  
CC cancer cells or treating a patient who bears cancer cells that expresses  
CC the protein, by administering the protein, antibody, polynucleotide,  
CC encoding the protein, antisense polynucleotide to the polynucleotide,  
CC ribozyme that cleaves the polynucleotide and T cells that specifically  
CC recognize the protein; and generating a mammalian immune response  
CC directed to the protein exposing cells of the mammal's immune system to  
CC an immunogenic portion of the protein or polynucleotide. The composition,  
CC which comprises an antibody specific to the protein, is useful for  
CC delivering a cytotoxic agent to a cell that expresses the protein by  
CC providing a cytotoxic agent conjugated to antibody and exposing the cell  
CC to the antibody-agent conjugate. The methods are useful for inhibiting



Query Match 9.4%; Score 113; DB 4; Length 213;  
 Best Local Similarity 27.1%; Pred. No. 0.00075;  
 Matches 52; Conservative 30; Mismatches 88; Indels 22; Gaps 8;

QY ASFTTSMEDTQVWVGSSPLGPAIGAEPPAGPOLPMLPERCAVFCQACIYVLA 90  
 DB ASMSMSEDAVS-----ADMERAGL--EEEAAB-----ERPLVFLCGGCRRLGDS 95  
 QY 91 VHLAMDLSR-SIGAVFRRVNNVLEAPFLVIGESLKGSTYNNLFCGSGGIPVGFELY 149  
 DB 96 --LSWVASQEDPTNLCILRCVSCNVSVDKEQKLKREKENGCVLFTLCCAGCSLNIYR 153  
 QY 150 STHALALRGHFCSSDKMVCYL--KTKAIVNASENDIQNVPLSEKIAELKEXIVLTH 207  
 DB 154 CTPKNLDYKRDILFCLSLVEAIESYVLGSSERKQIV--SEDKELFTL---ESRVELEKSLTQME 209  
 QY 208 NRLKSLMKILSE 219  
 DB 210 DVLKALQMKLWE 221

RESULT 7  
 ABB54167  
 ID ABB54167 standard; protein; 695 AA.

AC ABB54167;  
 XX 29-AUG-2003 (revised)  
 DT 16-MAY-2002 (first entry)  
 XX  
 DE Lactococcus lactis protein y1hc.  
 XX  
 XX Biosynthesis; biodegradation; lactic bacterium; yogurt; cheese.  
 XX  
 OS Lactococcus lactis; IL1403.

XX FR2807446-A1.  
 XX 12-OCT-2001.  
 PD  
 XX 11-APR-2000; 2000FR-00004630.  
 PF  
 XX 11-APR-2000; 2000FR-00004630.  
 PR

XX (INRG ) INRA INST NAT RECH AGRONOMIQUE.

XX Bolotine A, Sorokine A, Renault P, Ehrlich SD;

XX WPI; 2002-043418/06.

XX New nucleotide sequence useful in the identification or Lactococcus  
 PT lactis and related species.  
 XX

PS Claim 6; SEQ ID NO 869; 2504pp; French.

XX The present invention is related to a Lactococcus lactis nucleotide  
 CC sequence (ABB90521) and related proteins (ABB53300-ABB55621). The nucleic  
 CC acid sequence is useful in the detection and/or amplification of nucleic  
 CC acid sequence, particularly to identify Lactococcus lactis or related  
 CC species. The proteins of the invention are useful for the biosynthesis or  
 CC biodegradation of a composition of interest. The invention helps research  
 CC in lactic bacteria, particularly useful in the production of yogurt and  
 CC cheese. Note: The sequence data for this patent is based on equivalent  
 CC patent WO200177334 (published 18-OCT-2001) which is available in  
 CC electronic format directly from WIPO at  
 CC ftp.wipo.int/pub/published\_sequences. (Updated on 29-AUG-2003 to  
 CC standardise OS field)  
 XX

XX Sequence 695 AA;

Query Match 7.7%; Score 92; DB 5; Length 695;  
 Best Local Similarity 25.3%; Pred. No. 0.96;

Matches 38; Conservative 28; Mismatches 50; Indels 34; Gaps 7;

QY 101 LGAVFRRVNNVLEAPFLVIGESLKGSTYNNLFCGSGGIPVGFHLYSTHAAALARG 160  
 DB 101 LGTILF-----FVSGTFFSGARGELKRRPPAMMLITWGITVAY-AYSYATITMSLNG 153  
 QY 161 HF-----CLSSDKMVCYLKTKAIVNASE--MDI-----QNVPLSE-K 195  
 DB 154 HMGANFPELALTLYVIMLGHLEIKAIMGADALKDIALVPPKKAHLKSGDVELSELK 213  
 QY 196 IAELEKIVLTHNRLKSLMKILSEVTPDQS 225  
 DB 214 VGDL--LLVKNNEKIPADGLITSEALVDES 241

RESULT 8  
 AAO08972  
 ID AAO08972 standard; protein; 269 AA.

AC AAO08972;

XX 06-NOV-2001 (first entry)  
 DT  
 XX

DE Human polypeptide SEQ ID NO 22864.

XX Human; cytokine; cell proliferation; cell differentiation; gene therapy;  
 KM vaccine; peptide therapy; stem cell growth factor; haematopoiesis;  
 KM tissue growth factor; immunomodulatory; cancer; leukaemia;  
 KM nervous system disorders; arthritis; inflammation.  
 XX

OS Homo sapiens.

PN WO200164835-A2.

XX 07-SEP-2001.  
 PD

PF 26-FEB-2001; 2001WO-US004927.

XX 28-FEB-2000; 2000US-00515126.  
 PR 18-MAY-2000; 2000US-00577409.

XX (HYSE-) HYSEQ INC.

XX Tang YT, Liu C, Dymnac RT;

XX WPI; 2001-514838/56.

DR N-PSDB; AAI88903.

XX Isolated nucleic acids and polypeptides, useful for preventing diagnosing  
 PT and treating e.g. leukemia, inflammation and immune disorders.  
 XX

PS Claim 20; SEQ ID NO 22864; 1399pp + Sequence Listing; English.

XX The invention relates to human polynucleotides (AAI79941-AAI93841) and  
 CC the encoded proteins (AAO00010-AAO13910) that exhibit activity elating to  
 CC cytokine, cell proliferation or cell differentiation or which may induce  
 CC production of other cytokines in other cell populations. The  
 CC polynucleotides and polypeptides are useful in gene therapy, vaccines or  
 CC peptide therapy. The polypeptides have various cytokine-like activities,  
 CC e.g. stem cell growth factor activity, haematopoiesis regulating  
 CC activity, tissue growth factor activity, immunomodulatory activity and  
 CC activin/inhibin activity and may be useful in the diagnosis and/or  
 CC treatment of cancer, leukaemia, nervous system disorders, arthritis and  
 CC inflammation. Note: The sequence data for this patent did not form part  
 CC of the printed specification, but was obtained in electronic format  
 CC directly from WIPO at ftp.wipo.int/pub/published\_sequences

XX Sequence 269 AA;

Query Match 7.5%; Score 90; DB 4; Length 269;  
 Best Local Similarity 23.0%; Pred. No. 0.38;  
 Matches 52; Conservative 26; Mismatches 74; Indels 74; Gaps 11;

```

QY      8 HSRCATPPRGDFC-----GTERAIDA-----SFTSMEMDQVVKSSPL----- 50
Db      39 YRQAAQPPH-----CPAPEGPGAPQALGDAPSTVSILTAVQDYVCPQLGSHALCTCCPQ 95
QY      51 ---GPAGIGAEPPAAGPOLPSWLOPBCAVPQCAQCHAVLADSVHAMDLSR----- 99
Db      96 PMPDRAREEQDPVAPQ-----QCAVC---LQPFCHLYWGCTRTGCGGCLA 139
QY      100 ----SLGAVFSRVTNVNLVLEAPFLVIGESLKGSTY-NLFCGSCGIPVGFHLYSTHA 153
Db      140 PFCELNLGDKCLDGVANNNSYESDILKNYLAT-RGLTWKNML-----TES 183
QY      154 ALAALRGHFCLIS-----SDKWCYILKTKAIVNASEMDIQVPLSE 194
Db      184 LVALORGVFLISDYRVGTDTVLCYCCGLRSPRELTYQYQNIIPASE 229

RESULT 9
AAB93182
ID AAB93182 standard; protein; 623 AA.
XX
AC AAB93182;
XX
DE 26-JUN-2001 (first entry)
XX
DE Human protein sequence SEQ ID NO:12128.
XX
OS Homo sapiens.
XX
PN EPI074617-A2.
XX
PD 07-FEB-2001.
XX
PF 28-JUL-2000; 2000EP-00116126.
XX
PR 29-JUL-1999; 99JP-00248036.
PR 27-AUG-1999; 99DP-00300253.
PR 11-JAN-2000; 2000JP-00118776.
PR 02-MAY-2000; 2000JP-00183767.
PR 09-JUN-2000; 2000JP-00241899.
XX
PA (HELI-) HELIX RES INST.
XX
PI Ota T, Isegai T, Nishikawa T, Hayashi K, Saito K, Yamamoto J;
PI Ishii S, Sugiyama T, Wakamatsu A, Nagai K, Otsuki T;
XX
DR WPI; 2001-318749/34.
XX
PT Primer sets for synthesising polynucleotides, particularly the 5602 full-
PT length cDNAs defined in the specification, and for the detection and/or
PT diagnosis of the abnormality of the proteins encoded by the full-length
PT cDNAs.
XX
PS Claim 8; SEQ ID NO 12128; 2537PP + Sequence Listing; English.
XX
CC The present invention describes primer sets for synthesising 5602 full-
CC length cDNAs defined in the specification. Where a primer set comprises:
CC (a) an oligo-OT primer and an oligonucleotide complementary to the
CC complementary strand of a polynucleotide which comprises one of the 5602
CC nucleotide sequences defined in the specification; where the
CC oligonucleotide comprises at least 15 nucleotides; or (b) a combination
CC of an oligonucleotide comprising a sequence complementary to the
CC complementary strand of a polynucleotide which comprises a 5'-end
CC sequence and an oligonucleotide which comprises a 3'-end sequence, where the
CC oligonucleotide which comprises a 3'-end sequence, where the
CC oligonucleotide comprises at least 15 nucleotides and the combination of
CC the 5'-end sequence/3'-end sequence is selected from those defined in the
CC specification. The primer sets can be used in antisense therapy and in
CC gene therapy. The primers are useful for synthesising polynucleotides,
CC particularly full-length cDNAs. The primers are also useful for the
CC detection and/or diagnosis of the abnormality of the proteins encoded by

```

```

CC the full-length cDNAs. The primers allow obtaining of the full-length
CC cDNAs easily without any specialised methods. AAH03166 to AAH13628 and
CC AAH13633 to AAH18742 represent human cDNA sequences; AAB92446 to AAB95893
CC represent human amino acid sequences; and AAH13629 to AAH13632 represent
CC oligonucleotides, all of which are used in the exemplification of the
CC present invention
XX
SQ Sequence 623 AA;
XX
Query Match 74%; Score 89; DB 4; Length 623;
Best Local Similarity 23.0%; Pred. No. 1.8;
Matches 52; Conservative 26; Mismatches 74; Indels 74; Gaps 11;

QY      8 HSRCATPPRGDFC-----GTERAIDA-----SFTSMEMDQVVKSSPL----- 50
Db      39 YRQAAQPPH-----CPAPEGPGAPQALGDAPSTVSILTAVQDYVCPQLGSHALCTCCPQ 449
QY      51 ---GPAGIGAEPPAAGPOLPSWLOPBCAVPQCAQCHAVLADSVHAMDLSR----- 99
Db      450 PMPDRAREEQDPVAPQ-----QCAVC---LQPFCHLYWGCTRTGCGGCLA 493
QY      100 ----SLGAVFSRVTNVNLVLEAPFLVIGESLKGSTY-NLFCGSCGIPVGFHLYSTHA 153
Db      494 PFCELNLGDKCLDGVANNNSYESDILKNYLAT-RGLTWKNML-----TES 537
QY      154 ALAALRGHFCLIS-----SDKWCYILKTKAIVNASEMDIQVPLSE 194
Db      538 LVALORGVFLISDYRVGTDTVLCYCCGLRSPRELTYQYQNIIPASE 583

RESULT 10
AAB97233
ID AAB97233 standard; protein; 623 AA.
XX
AC AAB97233;
XX
DE 27-JUN-2002 (first entry)
XX
DE Novel human protein SEQ ID NO: 501.
XX
OS Homo sapiens.
XX
PN WO200222660-A2.
XX
PD 21-MAR-2002.
XX
PF 10-SEP-2001; 2001WO-US026015.
XX
PR 11-SEP-2000; 2000US-00659671.
XX
PA (HYSE-) HYSEQ INC.
XX
PI Tang YF, Liu C, Zhou P, Asundi V, Zhang J, Zhao QA, Ren F;
PI Xue AJ, Yang Y, Wehrman T, Drmanac RT;
XX
DR WPI; 2002-292408/33.
XX
DR N-PsDB; ABB32419.
XX
PT An isolated polynucleotide for treating diseases associated with its
PT encoded polypeptide such as cancer and multiple sclerosis.
XX
PS Example 2; SEQ ID NO 501; 509PP; English.
XX
CC The present invention provides the protein and coding sequences of 444
CC novel human proteins. These were isolated from expressed sequences tags
CC (ESTs). They can be used to stimulate cell growth, to regulate
CC haematopoiesis e.g. to treat aplastic anaemia, to help tissue regrowth
CC e.g. in burn treatment, to regulate the immune system e.g. to treat

```

CC multiple sclerosis, to regulate activin or inhibin e.g. to treat  
 CC infertility, to regulate haemostasis or thrombolysis e.g. to treat stroke  
 CC and cancer, to screen for drugs, to treat inflammatory conditions e.g.  
 CC rheumatoid arthritis, and to treat nervous system disorders e.g.  
 CC Parkinson's disease. The present sequence is a protein of the invention  
 SQ Sequence 623 AA:

Query Match 7.4%; Score 89; DB 5; Length 623;  
 Best Local Similarity 23.0%; Pred. No. 1.8;  
 Matches 52; Conservative 26; Mismatches 74; Indels 74; Gaps 11;

QY 8 HRSRCATPRGDFC-----GTERAIDQA-----SFTTSMEDTQVVKSSPL----- 50  
 DB 393 YRRQAAQPPH---CPAPGEPGAPQALGDAPPTSVSLTTAVQDYCPLOGSHALCTCCFQ 449  
 QY 51 --GPAGAGAEPPAQPQLPSWLQPERCAVFOCAQCHAVLADSVHLANDLSR----- 99  
 DB 450 PMPDRRAREDDPRVAPQ-----QCAYC---LQPFCHLYWGCTRTGCGYGLA 493  
 QY 100 -----SLGAVVFSRYTNVNLAPFLVIGESLKGSTY-NLIFCGSCGIPVGFHLYSTHA 153  
 DB 494 PFCEINTLGDCKLDGVLNNNSYESDLTKNYLAT-RGLTWKML-----TBS 537  
 QY 154 ALAALRGHFCLS-----SDKWCYLLKTKATAYNASMDIONVPLSE 194  
 DB 538 LVVALRGVFLSDYRVGTDTVLCYCGGLSRFRELTYQYQONIPASE 583

## RESULT 11

AAB93168  
 ID AAB93168 standard; protein; 652 AA.

AC AAB93168;  
 XX  
 DT 26-JUN-2001 (first entry)  
 XX  
 DE Human protein sequence SEQ ID NO:12100.  
 XX  
 KW Human; primer; detection; diagnosis; antisense therapy; gene therapy.  
 XX  
 OS Homo sapiens.  
 XX  
 PN EP1074617-A2.  
 XX  
 PD 07-FEB-2001.  
 XX  
 PF 28-JUL-2000; 2000EP-00116126.  
 XX  
 PR 29-JUL-1999; 99JP-00248036.  
 PR 27-AUG-1999; 99JP-00300253.  
 PR 11-JAN-2000; 2000JP-00118176.  
 PR 02-MAY-2000; 2000JP-00183767.  
 PR 09-JUN-2000; 2000JP-00241899.  
 XX

PA (HELI-) HELIX RES INST.  
 XX  
 PI Ota T, Isogai T, Nishikawa T, Hayashi K, Saito K, Yamamoto J;  
 PI Ishii S, Sugiyama T, Wakamatsu A, Nagai K, Otsuki T;  
 XX  
 DR WPI; 2001-318749/34.  
 XX  
 XX

PT Primer sets for synthesizing polynucleotides, particularly the 5602 full-  
 PT length cDNAs defined in the specification, and for the detection and/or  
 PT diagnosis of the abnormality of the proteins encoded by the full-length  
 PT cDNAs.  
 XX

PS Claim 8; SEQ ID NO 12100; 2537bp + Sequence listing; English.  
 XX

CC The present invention describes primer sets for synthesizing 5602 full-  
 CC length cDNAs defined in the specification. Where a primer set comprises:  
 CC (a) an oligo-dT primer and an oligonucleotide complementary to the  
 CC complementary strand of a polynucleotide which comprises one of the 5602

CC nucleotide sequences defined in the specification, where the  
 CC oligonucleotide comprises at least 15 nucleotides; or (b) a combination  
 CC of an oligonucleotide comprising a sequence complementary to the  
 CC complementary strand of a polynucleotide which comprises a 5'-end  
 CC sequence and an oligonucleotide comprising a sequence complementary to a  
 CC polynucleotide which comprises a 3'-end sequence, where the  
 CC oligonucleotide comprises at least 15 nucleotides and the combination of  
 CC the 5'-end sequence/3'-end sequence is selected from those defined in the  
 CC specification. The primer sets can be used in antisense therapy and in  
 CC gene therapy. The primers are useful for synthesizing polynucleotides,  
 CC particularly full-length cDNAs. The primers are also useful for the  
 CC detection and/or diagnosis of the abnormality of the proteins encoded by  
 CC the full-length cDNAs. The primers allow obtaining of the full-length  
 CC cDNAs easily without any specialised methods. AAB93166 to AAB13628 and  
 CC AAB13633 to AAB18742 represent human cDNA sequences; AAB92446 to AAB95893  
 CC represent human amino acid sequences; and AAB13629 to AAB13632 represent  
 CC oligonucleotides, all of which are used in the exemplification of the  
 CC present invention  
 SQ Sequence 652 AA:

Query Match 7.4%; Score 89; DB 4; Length 652;  
 Best Local Similarity 23.0%; Pred. No. 1.9;  
 Matches 52; Conservative 26; Mismatches 74; Indels 74; Gaps 11;

QY 8 HRSRCATPRGDFC-----GTERAIDQA-----SFTTSMEDTQVVKSSPL----- 50  
 DB 422 YRRQAAQPPH---CPAPGEPGAPQALGDAPPTSVSLTTAVQDYCPLOGSHALCTCCFQ 478  
 QY 51 --GPAGAGAEPPAQPQLPSWLQPERCAVFOCAQCHAVLADSVHLANDLSR----- 99  
 DB 479 PMPDRRAREDDPRVAPQ-----QCAYC---LQPFCHLYWGCTRTGCGYGLA 522  
 QY 100 -----SLGAVVFSRYTNVNLAPFLVIGESLKGSTY-NLIFCGSCGIPVGFHLYSTHA 153  
 DB 523 PFCEINTLGDCKLDGVLNNNSYESDLTKNYLAT-RGLTWKML-----TBS 566  
 QY 154 ALAALRGHFCLS-----SDKWCYLLKTKATAYNASMDIONVPLSE 194  
 DB 567 LVVALRGVFLSDYRVGTDTVLCYCGGLSRFRELTYQYQONIPASE 612

## RESULT 12

AAB83843  
 ID AAB83843 standard; protein; 664 AA.

AC AAB83843;  
 XX  
 DT 22-AUG-2001 (first entry)  
 XX  
 DE Amino acid sequence of a human ring finger protein designated FHAR1.  
 XX  
 KW FHAR1; RING finger protein; cancer; vaccine.  
 XX  
 OS Homo sapiens.  
 XX  
 PN WO200142430-A1.  
 XX  
 PD 14-JUN-2001.  
 XX  
 PF 07-DEC-2000; 2000WO-US033094.  
 XX  
 PR 08-DEC-1999; 99US-00456876.  
 PR (SMIK ) SMITHKLINE BEECHAM CORP.  
 PA  
 PI Zhou B, Zhu Y, Chaturvedi P, Hurler MR, Li X;  
 XX  
 DR WPI; 2001-381663/40.  
 DR N-PSDB; AAF89709.  
 XX

PT New FHAR1 polypeptide, a member of the RING finger protein family for  
 PT diagnosing and treating cancer, and for use in anti-cancer vaccines.



KM antiulcer; osteopathic; immunosuppressive; antiinflammatory; cyostatic;  
 KM gene therapy; chromosome 12.  
 XX Homo sapiens.  
 OS WO2003029271-A2.  
 XX 10-APR-2003.  
 PD 24-SEP-2002; 2002WO-US030474.  
 XX 24-SEP-2001; 2001US-0324631P.  
 PR (HYSE-) HYSEQ INC.  
 XX  
 PA Tang TY, Zhang J, Ren F, Xue AJ, Zhao QA, Wang J, Wehrman T;  
 PI Zhou P, Ghosh M, Wang D, Asundi V, Wang Z, Weng G;  
 PI Haley-Vicente D, Drmanac RT;  
 XX  
 DR MPI; 2003-371981/35.  
 DR N-PSDB; ADC32051.  
 XX  
 XX New polynucleotide and polypeptide useful for diagnosing, preventing or  
 PT treating conditions such as neurodegenerative diseases, anemias, platelet  
 PT disorders, wounds, burns, ulcers, osteoporosis, autoimmune diseases or  
 PT cancer.  
 XX  
 PS Example 2; SEQ ID NO 2900; 1185bp; English.  
 XX  
 XX The invention relates to 971 novel human cDNA sequences (ADC29919-  
 CC ADC30889) and the polypeptides they encode (ADC30890-ADC31860). The  
 CC invention also relates to nucleic acid sequences over 99% identical with  
 CC the novel human cDNAs. The invention additionally encompasses expression  
 CC vectors and host cells comprising a nucleic acid of the invention; the  
 CC recombinant production of a polypeptide of the invention; an antibody  
 CC against a polypeptide of the invention; a method of detecting  
 CC polynucleotides or polypeptides of the invention; and methods of  
 CC identifying a compound which binds to a polypeptide of the invention. The  
 CC invention further discloses methods of preventing, treating or  
 CC ameliorating a medical condition; kits comprising polynucleotide probes  
 CC and/or monoclonal antibodies for carrying out the methods of the  
 CC invention; methods for the identification of compounds that modulate the  
 CC expression or activity of the polynucleotide and/or polypeptide; and 767  
 CC contig sequences corresponding to the cDNA sequences of the invention  
 CC (ADC31861-ADC32627) and the polypeptides encoded by the contigs (ADC32628  
 CC -ADC33394). The nucleic acids and polypeptides of the invention are  
 CC useful in diagnostics, drug screening, forensics, gene mapping, in the  
 CC identification of mutations responsible for genetic disorders or other  
 CC traits, for assessing biodiversity, and in producing many other types of  
 CC data and products dependent on DNA and amino acid sequences. They are  
 CC also used for treating diseases such as Parkinson's disease, Alzheimer's  
 CC disease and other neurodegenerative diseases, anaemia, platelet  
 CC disorders, wounds, burns, ulcers, osteoporosis, autoimmune diseases or  
 CC cancer. The nucleic acids may also be used as hybridisation probes or  
 CC primers, and in the recombinant production of a protein. The polypeptides  
 CC are also useful in generating antibodies, as molecular weight markers,  
 CC and as food supplements. The present sequence represents a human contig-  
 CC encoded polypeptide sequence used in an example of the invention. Note:  
 CC The sequence data for this patent did not form part of the printed  
 CC specification, but was obtained in electronic format directly from WIPO  
 CC at ftp.wipo.int/pub/published\_pct\_sequences.  
 XX  
 SQ Sequence 1245 AA;  
 Query Match 7.3%; Score 87.5; DB 7; Length 1245;  
 Best Local Similarity 21.8%; Pred. No. 7.6;  
 Matches 51; Conservative 35; Mismatches 107; Indels 41; Gaps 7;  
 QY 25 ERAIDQASFTTSMWDQVQVKGSSPLGPAGIGAEPAAGPOL-----PSW--LQPERCAV 77  
 DB 610 ERHYRDRIETRYGTSTRIQDLPDPGRKREKGPPLPSGSGKAVASSPRHENTPRRAAI 669  
 QY 78 FOCAQCHAVLAD--SVHLAMDLSRSLGAVVFSRVTNVNLAEPLVIGISLKGSTYNTL 134

DB 670 VLTSALGGGLASQCGGLPWSWCLQGFIMAKYPLKPRSKULTUMPVISIAAGR-ENSKRL 728  
 QY 135 LFCGSCGIP-----VGFHLYSTHAALALRGHPLCSDRKWCYLAK 175  
 DB 729 QKFGSLGFPPSSRRTRFTYRTTGPFPLVNHICTHSDCASWGVCEVSSQDBQHSQY 788  
 QY 176 TKAIVNASEMDIQNVPLSEKIA---ELKEK-----YLTNRKSLMKILS 218  
 DB 789 SKLLASKEQVLRHVLEEKVLAECQLAEKHTPESCFEALIOELKGVLELYLS 842  
 RESULT 15  
 ABU23878  
 ID ABU23878 standard; protein; 669 AA.  
 XX  
 AC ABU23878;  
 XX  
 DT 19-JUN-2003 (first entry)  
 XX  
 DE Protein encoded by Prokaryotic essential gene #9405.  
 XX  
 KM Antisense; prokaryotic essential gene; cell proliferation; drug design.  
 XX  
 OS Clostridium acetobutylicum.  
 XX  
 PN WO200277183-A2.  
 XX  
 PD 03-OCT-2002.  
 XX  
 XX 21-MAR-2002; 2002WO-US009107.  
 PF  
 XX 21-MAR-2001; 2001US-00815242.  
 PR 06-SEP-2001; 2001US-00948993.  
 PR 25-OCT-2001; 2001US-0342923P.  
 PR 08-FEB-2002; 2002US-00072851.  
 PR 06-MAR-2002; 2002US-0362699P.  
 XX  
 PA (ELIT-) ELITRA PHARM INC.  
 XX  
 XX Wang L, Zamudio C, Malone C, Haseelbeck R, Ohlsen KL, Zykkind JW;  
 PI Wall D, Trawick JD, Carr GJ, Yamamoto R, Forsyth RA, Xu HH;  
 XX  
 DR MPI; 2003-029926/02.  
 DR N-PSDB; ACA27748.  
 XX  
 PS New antisense nucleic acids, useful for identifying proteins or screening  
 PT for homologous nucleic acids required for cellular proliferation to  
 PT isolate candidate molecules for rational drug discovery programs.  
 XX  
 XX Claim 25; SEQ ID NO 51802; 1766bp; English.  
 XX  
 CC The invention relates to an isolated nucleic acid comprising any one of  
 CC the 6213 antisense sequences given in the specification where expression  
 CC of the nucleic acid inhibits proliferation of a cell. Also included are:  
 CC (1) a vector comprising a promoter operably linked to the nucleic acid  
 CC encoding a polypeptide whose expression is inhibited by the antisense  
 CC nucleic acid; (2) a host cell containing the vector; (3) an isolated  
 CC polypeptide or its fragment whose expression is inhibited by the  
 CC antisense nucleic acid; (4) an antibody capable of specifically binding  
 CC the polypeptide; (5) producing the polypeptide; (6) inhibiting cellular  
 CC proliferation or the activity of a gene in an operon required for  
 CC proliferation; (7) identifying a compound that influences the activity of  
 CC the gene product or that has an activity against a biological pathway;  
 CC required for proliferation, or that inhibits cellular proliferation; (8)  
 CC identifying a gene required for cellular proliferation or the biological  
 CC pathway in which a proliferation-required gene or its gene product lies  
 CC or a gene on which the test compound that inhibits proliferation of an  
 CC organism acts; (9) manufacturing an antibiotic; (10) profiling a  
 CC compound's activity; (11) a culture comprising strains in which the gene  
 CC product is overexpressed or underexpressed; (12) determining the extent  
 CC to which each of the strains is present in a culture or collection of  
 CC strains; or (13) identifying the target of a compound that inhibits the

CC proliferation of an organism. The antisense nucleic acids are useful for  
CC identifying proteins or screening for homologous nucleic acids required  
CC for cellular proliferation to isolate candidate molecules for rational  
CC drug discovery programs, or for screening homologous nucleic acids  
CC required for proliferation in cells other than *S. aureus*, *S. typhimurium*,  
CC *K. pneumoniae* or *P. aeruginosa*. The present sequence is encoded by one of  
CC the target prokaryotic essential genes. Note: The sequence data for this  
CC patent did not form part of the printed specification, but was obtained  
CC in electronic format directly from WIPO at  
CC ftp.wipo.int/pub/published\_pct\_sequences  
XX

Sequence 669 AA;

Query Match 7.2%; Score 86.5; DB 6; Length 669;

Best Local Similarity 27.4%; Pred. No. 3.8;

Matches 37; Conservative 22; Mismatches 45; Indels 31; Gaps 8;

```
QY 96 DLSR--SLGAVFSRVTVNNVLEAPFLVGIKSTYNTL-----LPGSCGIPVGF 146
   | | | | | | | | | | | | | | | | | | | | | | | | | | | |
Db 366 DIGKKVKYIGSRVFRSRNDVI--PRIMGVTEFEGTNEIEAPTICPGSEIVKEGV 422
   | | | | | | | | | | | | | | | | | | | | | | | | | | | |
QY 147 HLYSTHAALALRGHFLSSDKWVCYLKTKATYNASEMDIQNP-LSEKIAR-LKEKIV 204
   | | | | | | | | | | | | | | | | | | | | | | | | | | | |
Db 423 HL-----FC-ENTLSCKPQWYKSIHFASREAMNIEGPFSEKTAEOLEFEK-- 465
   | | | | | | | | | | | | | | | | | | | | | | | | | | | |
QY 205 LTHNRKLSMKILSE 219
   | | | | | | | | | | | | | | | | | | | | | | | | | | | |
Db 466 LNIKISIDLYRITKE 480
```

Search completed: July 20, 2004, 11:16:29  
Job time : 56 secs

GenCore version 5.1.6  
Copyright (c) 1993 - 2004 Compugen Ltd.

## OM protein - protein search, using sw model

Run on: July 20, 2004, 11:05:13 ; Search time 19 Seconds

(without alignments)  
622.229 Million cell updates/sec

Title: US-09-942-052A-728

Perfect score: 1198  
Sequence: 1 MAAQPLRHRSRCATPPRGDF.....LKSIMKILSEVTPDQSPEN 229Scoring table: BLOSUM62  
Gapop 10.0 , Gapext 0.5

Searched: 389414 seqs, 51625971 residues

Total number of hits satisfying chosen parameters: 389414

Minimum DB seq length: 0  
Maximum DB seq length: 2000000000Post-processing: Minimum Match 0%  
Maximum Match 100%  
Listing first 45 summaries

## Database :

Issued Patents AA:\*  
1: /cgn2\_6/prodata/2/1aa/5A.COMB.pep:\*  
2: /cgn2\_6/prodata/2/1aa/5B.COMB.pep:\*  
3: /cgn2\_6/prodata/2/1aa/6A.COMB.pep:\*  
4: /cgn2\_6/prodata/2/1aa/6B.COMB.pep:\*  
5: /cgn2\_6/prodata/2/1aa/PCTUS.COMB.pep:\*  
6: /cgn2\_6/prodata/2/1aa/backfile1.pep:\*

Pred. No. is the number of results predicted by chance to have a  
score greater than or equal to the score of the result being printed,  
and is derived by analysis of the total score distribution.

## SUMMARIES

Result No.	Score	Query Match Length	ID	Description
1	86	7.2	158 4 US-09-134-000C-6596	Sequence 6596, Ap
2	83	6.9	457 6 5268463-7	Patent No. 5268463
3	81.5	6.8	456 6 5432081-7	Patent No. 5432081
4	81.5	6.8	457 2 US-08-882-704A-6	Sequence 6, Appl1
5	81.5	6.8	457 4 US-09-151-957-6	Sequence 6, Appl1
6	81.5	6.8	457 4 US-10-195-158-6	Sequence 6, Appl1
7	80.5	6.7	1711 2 US-08-342-930-2	Sequence 2, Appl1
8	78.5	6.6	325 2 US-09-018-576-12	Sequence 3, Appl1
9	78.5	6.6	325 3 US-09-248-137-12	Sequence 12, Appl1
10	78.5	6.6	325 3 US-09-248-137-12	Sequence 3, Appl1
11	78.5	6.6	325 3 US-09-248-137-12	Sequence 12, Appl1
12	78.5	6.6	325 3 US-09-248-137-12	Sequence 3, Appl1
13	77	6.4	607 4 US-09-837-682-1	Sequence 1, Appl1
14	77	6.4	719 2 US-08-520-933-3	Sequence 3, Appl1
15	77	6.4	719 4 US-09-285-040-3	Sequence 3, Appl1
16	77	6.4	738 6 5262177-2	Patent No. 5262177
17	76.5	6.3	310 1 US-08-845-713A-2	Sequence 2, Appl1
18	75.5	6.3	310 1 US-08-129-456A-36	Sequence 36, Appl1
19	75.5	6.3	310 2 US-08-705-868-3	Sequence 3, Appl1
20	75.5	6.3	310 3 US-09-123-615-3	Sequence 3, Appl1
21	75.5	6.3	310 3 US-08-360-821B-35	Sequence 35, Appl1
22	75.5	6.3	1449 3 US-08-840-062-6	Sequence 6, Appl1
23	75.5	6.3	1723 4 US-09-194-612A-31	Sequence 31, Appl1
24	75.5	6.3	4545 2 US-08-804-227C-14	Sequence 14, Appl1
25	75.5	6.3	4550 2 US-08-804-227C-8	Sequence 8, Appl1
26	75.5	6.3	4550 2 US-08-804-198-2	Sequence 2, Appl1
27	75	6.3	458 4 US-09-800-170-3	Sequence 3, Appl1

28	75	6.3	717 6 5262177-5	Patent No. 5262177
29	75	6.3	1088 4 US-09-920-804-2	Sequence 2, Appl1
30	75	6.3	1088 4 US-09-920-804-4	Sequence 4, Appl1
31	75	6.3	1129 4 US-09-734-674-2	Sequence 2, Appl1
32	75	6.3	2227 3 US-08-475-886-4	Sequence 4, Appl1
33	75	6.3	2227 4 US-09-653-499-4	Sequence 4, Appl1
34	75	6.3	2227 4 US-10-135-988-4	Sequence 4, Appl1
35	74	6.2	380 3 US-08-801-344-11	Sequence 11, Appl1
36	74	6.2	380 4 US-09-498-599-11	Sequence 11, Appl1
37	74	6.2	2109 3 US-08-646-695-6	Sequence 6, Appl1
38	74	6.2	2109 5 PCT-US96-06053-6	Sequence 6, Appl1
39	74	6.2	2227 3 US-08-475-886-6	Sequence 6, Appl1
40	74	6.2	2227 3 US-08-397-232-4	Sequence 4, Appl1
41	74	6.2	2227 4 US-09-653-489-6	Sequence 6, Appl1
42	74	6.2	2227 4 US-10-135-988-6	Sequence 6, Appl1
43	73.5	6.1	589 2 US-08-453-848-13	Sequence 13, Appl1
44	73.5	6.1	589 4 US-09-169-027-13	Sequence 13, Appl1
45	73.5	6.1	803 4 US-09-786-240-12	Sequence 12, Appl1

## ALIGNMENTS

```
RESULT 1
US-09-134-000C-6596
; Sequence 6596, Application US/09134000C
; Patent No. 6617156
; GENERAL INFORMATION:
; APPLICANT: Lynn Doucette-Stamm et al
; TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO
; FILE REFERENCE: 032796-032
; CURRENT APPLICATION NUMBER: US/09/134,000C
; PRIOR FILING DATE: 1998-08-13
; PRIOR APPLICATION NUMBER: US 60/055,778
; NUMBER OF SEQ ID NOS: 6812
; SOFTWARE: PatentIn version 3.1
; SEQ ID NO 6596
; LENGTH: 158
; TYPE: PRT
; ORGANISM: Enterococcus faecalis
US-09-134-000C-6596

Query Match
Best Local Similarity 7.2%; Score 86; DB 4; Length 158;
Matches 40; Conservative 27; Mismatches 54; Indels 30; Gaps 9;

QY 87 LADSVHLAMDLSRSIGAVY---FSRYTNVVLEA--PLVIGISGSIKG-----STYN 133
|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|
Db 17 LVEYTLASEOPLSTGSTITLNGSLKPTGNDLNKAFFFIINDEGDPGKFSYGNLYKTIN 76
QY 134 LIFGSGCGIPVGPHLYSTHAALAL--RGHPTCLSDKVCYLLKTAIVNASEMDIQNY 150
|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|
Db 77 YQVCNSGCSLPDSQNIETSSALALGYIRNH-----KNTC---TKLYTILSLNGFENK 128
QY 191 PLSE---KIAEKIKYIVLTNFKLSIMKL 217
|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|
Db 129 PISKTVRKLSIDSYK-DIYYQEMK-ILKII 157

RESULT 2
5268463-7
; Patent No. 5268463
; APPLICANT: JEFFERSON, RICHARD A.
; TITLE OF INVENTION: PLAAAT PROMOTER a-GLUCURONIDASE GENE
; CONSTRUCT
; NUMBER OF SEQUENCES: 9
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/07/447,976
; FILING DATE: 08-DEC-1989
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: 119,102
```

FILED DATE: 10-NOV-1987  
APPLICATION NUMBER: 264,586  
FILING DATE: 31-OCT-1986  
SEQ ID NO: 7  
LENGTH: 457  
5268463-7

Query Match 6.9%; Score 83; DB 6; Length 457;  
Best Local Similarity 25.2%; Pred. No. 1.3;  
Matches 37; Conservative 30; Mismatches 54; Indels 26; Gaps 8;

QY 95 WDLRSLSGAIVFSRYTNNV-----VLEA-----PFLVIGESLKGSTYNNL-FCGSCG 141  
DB 316 WSLPVALVALAIASIGQVMTVMALADTVYEGYLTGVR--IEGLTYSLSFSTRKCG 373  
QY 142 IPVGFHLVSTHAALALRGHFC--LSSDKWVCYLTKKAIYVASEMDIQNV-----PLSE 194  
DB 374 QAIIGS--SIPAFILGSGYIANOVQTPREVIMGIRTSIALVPCGFMFLAFVIIMFYPLTD 431  
QY 195 KIAELKEKIVLTHNRLKSLMKILSEVT 221  
DB 432 K--KFKETVEIDNRKKVQQQLISDIT 456

RESULT 3  
5432081-7  
PATENT No. 5432081  
APPLICANT: JEFFERSON, RICHARD A.  
TITLE OF INVENTION: HOST CELLS TRANSFORMED WITH THE E. COLI  
GLUCONONIDE PERMEASE GENE  
NUMBER OF SEQUENCES: 10  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/08/138,546  
FILING DATE: 15-OCT-1993  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: 447,976  
FILING DATE: 08-DEC-1989  
APPLICATION NUMBER: 264,586  
FILING DATE: 31-OCT-1986  
APPLICATION NUMBER: 119,102  
FILING DATE: 10-NOV-1987  
SEQ ID NO: 7  
LENGTH: 456  
5432081-7

Query Match 6.8%; Score 81.5; DB 6; Length 456;  
Best Local Similarity 25.2%; Pred. No. 1.9;  
Matches 37; Conservative 30; Mismatches 53; Indels 27; Gaps 8;

QY 95 WDLRSLSGAIVFSRYTNNV-----VLEA-----PFLVIGESLKGSTYNNL-FCGSCG 141  
DB 316 WSLPVALVALAIASIGQVMTVMALADTVYEGYLTGVR--IEGLTYSLSFSTRKCG 373  
QY 142 IPVGFHLVSTHAALALRGHFC--LSSDKWVCYLTKKAIYVASEMDIQNV-----PLSE 194  
DB 374 QAIIGS--SIPAFILGSGYIANOVQTPREVIMGIRTSIALVPCGFMFLAFVIIMFYPLTD 430  
QY 195 KIAELKEKIVLTHNRLKSLMKILSEVT 221  
DB 431 K--KFKETVEIDNRKKVQQQLISDIT 455

RESULT 4  
US-08-882-704A-6  
Sequence 6, Application US/08882704A  
Patent No. 5879906  
GENERAL INFORMATION:  
APPLICANT: Jefferson, Richard A.  
APPLICANT: Wilson, Katherine J.  
APPLICANT: Leader, Michael  
TITLE OF INVENTION: GLUCONONIDE REPRESSORS AND USES THEREOF  
NUMBER OF SEQUENCES: 19  
CORRESPONDENCE ADDRESS:

ADDRESSEE: SEED and BERRY LLP  
STREET: 6300 Columbia Center, 701 Fifth Avenue  
CITY: Seattle  
STATE: Washington  
COUNTRY: USA  
ZIP: 98104-7092  
COMPUTER READABLE FORM:  
MEDIUM TYPE: Floppy disk  
COMPUTER: IBM PC compatible  
OPERATING SYSTEM: PC-DOS/MS-DOS  
SOFTWARE: Patent In Release #1.0, Version #1.30  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/08/882,704A  
FILING DATE: 25-JUN-1997  
CLASSIFICATION: 435  
ATTORNEY/AGENT INFORMATION:  
NAME: No. 5879906tenburg Ph.D., Carol  
REGISTRATION NUMBER: 39,317  
REFERENCE/DOCKET NUMBER: 190106.404  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: (206) 622-4900  
TELEFAX: (206) 682-6031  
INFORMATION FOR SEQ ID NO: 6:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 457 amino acids  
TYPE: amino acid  
STRANDEDNESS: single  
TOPOLOGY: linear  
US-08-882-704A-6

Query Match 6.8%; Score 81.5; DB 2; Length 457;  
Best Local Similarity 25.2%; Pred. No. 1.9;  
Matches 37; Conservative 30; Mismatches 53; Indels 27; Gaps 8;

QY 95 WDLRSLSGAIVFSRYTNNV-----VLEA-----PFLVIGESLKGSTYNNL-FCGSCG 141  
DB 317 WSLPVALVALAIASIGQVMTVMALADTVYEGYLTGVR--IEGLTYSLSFSTRKCG 374  
QY 142 IPVGFHLVSTHAALALRGHFC--LSSDKWVCYLTKKAIYVASEMDIQNV-----PLSE 194  
DB 375 QAIIGS--SIPAFILGSGYIANOVQTPREVIMGIRTSIALVPCGFMFLAFVIIMFYPLTD 431  
QY 195 KIAELKEKIVLTHNRLKSLMKILSEVT 221  
DB 432 K--KFKETVEIDNRKKVQQQLISDIT 456

RESULT 5  
US-09-151-957-6  
Sequence 6, Application US/09151957  
Patent No. 6429292  
GENERAL INFORMATION:  
APPLICANT: Jefferson, Richard A.  
APPLICANT: Wilson, Katherine J.  
APPLICANT: Leader, Michael  
TITLE OF INVENTION: GLUCONONIDE REPRESSORS AND USES THEREOF  
NUMBER OF SEQUENCES: 19  
CORRESPONDENCE ADDRESS:  
ADDRESSEE: SEED and BERRY LLP  
STREET: 6300 Columbia Center, 701 Fifth Avenue  
CITY: Seattle  
STATE: Washington  
COUNTRY: USA  
ZIP: 98104-7092  
COMPUTER READABLE FORM:  
MEDIUM TYPE: Floppy disk  
COMPUTER: IBM PC compatible  
OPERATING SYSTEM: PC-DOS/MS-DOS  
SOFTWARE: Patent In Release #1.0, Version #1.30  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/09/151,957  
FILING DATE: 11-SEP-1998  
CLASSIFICATION: <Unknown>

PRIOR APPLICATION DATA:  
APPLICATION NUMBER: US 08/882,704  
FILING DATE: <Unknown>  
ATTORNEY/AGENT INFORMATION:  
NAME: No. 6423292lemburg Ph.D., Carol  
REGISTRATION NUMBER: 39,317  
REFERENCE/DOCKET NUMBER: 190106.404  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: (206) 622-4900  
TELEFAX: (206) 682-6031  
INFORMATION FOR SEQ ID NO: 6:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 457 amino acids  
TYPE: amino acid  
STRANDEDNESS: single  
TOPOLOGY: linear  
SEQUENCE DESCRIPTION: SEQ ID NO: 6:  
US-09-151-957-6

Query Match 6.8%; Score 81.5; DB 4; Length 457;  
Best Local Similarity 25.2%; Pred. No. 1.9; Mismatches 53; Indels 27; Gaps 8;  
Matches 37; Conservative 30; Mismatches 53; Indels 27; Gaps 8;  
QY 95 WDLRSIGAVVFSRYTNV-----VLEA-----PFLVIGESLKSSTYML-FCGSCG 141  
317 WSLPVALVALAIASIGGVTTVMWALEADTVEGEYLTGVR--IEGLTYSLSFSTRKCG 374  
DB 142 IPVGPHLYSTHAALALRGHFC--LSSDKMVCYLKTKAIYNASEMDIQNV-----PLSE 194  
QY 375 QAIG--GSIPAFILIGSGYIANOVQTPVIMGIRTSIALVPCGFMILAFYIMYPPLTD 431  
DB 195 KIAELKEKIVLTHNRLKSLMKILSEVT 221  
QY 432 K-KFEKIVEIDNRKKVQQQLISDIT 456  
DB

RESULT 6  
US-10-195-158-6  
Sequence 6, Application US/10195158  
Patent No. 6659764  
GENERAL INFORMATION:  
APPLICANT: Jefferson, Richard A.  
Wilson, Katherine J.  
Leader, Michael  
TITLE OF INVENTION: GLUTCHRONIDE REPRESSORS AND USES THEREOF  
NUMBER OF SEQUENCES: 19  
CORRESPONDENCE ADDRESS:  
ADDRESSEE: SEED and BERRY LLP  
STREET: 6300 Columbia Center, 701 Fifth Avenue  
City: Seattle  
STATE: Washington  
COUNTRY: USA  
ZIP: 98104-7092  
COMPUTER READABLE FORM:  
MEDIUM TYPE: Floppy disk  
COMPUTER: IBM PC compatible  
OPERATING SYSTEM: PC-DOS/MS-DOS  
SOFTWARE: Patentin Release #1.0, Version #1.30  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/10/195,158  
FILING DATE: 08-Aug-2002  
CLASSIFICATION: <Unknown>  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: US/09/151,957  
FILING DATE: 11-Sep-1998  
APPLICATION NUMBER: US 08/882,704  
FILING DATE: <Unknown>  
ATTORNEY/AGENT INFORMATION:  
NAME: No. 6659764lemburg Ph.D., Carol  
REGISTRATION NUMBER: 39,317  
REFERENCE/DOCKET NUMBER: 190106.404  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: (206) 622-4900

TELEFAX: (206) 682-6031  
INFORMATION FOR SEQ ID NO: 6:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 457 amino acids  
TYPE: amino acid  
STRANDEDNESS: single  
TOPOLOGY: linear  
SEQUENCE DESCRIPTION: SEQ ID NO: 6:  
US-10-195-158-6

Query Match 6.8%; Score 81.5; DB 4; Length 457;  
Best Local Similarity 25.2%; Pred. No. 1.9; Mismatches 53; Indels 27; Gaps 8;  
Matches 37; Conservative 30; Mismatches 53; Indels 27; Gaps 8;  
QY 95 WDLRSIGAVVFSRYTNV-----VLEA-----PFLVIGESLKSSTYML-FCGSCG 141  
317 WSLPVALVALAIASIGGVTTVMWALEADTVEGEYLTGVR--IEGLTYSLSFSTRKCG 374  
DB 142 IPVGPHLYSTHAALALRGHFC--LSSDKMVCYLKTKAIYNASEMDIQNV-----PLSE 194  
QY 375 QAIG--GSIPAFILIGSGYIANOVQTPVIMGIRTSIALVPCGFMILAFYIMYPPLTD 431  
DB 195 KIAELKEKIVLTHNRLKSLMKILSEVT 221  
QY 432 K-KFEKIVEIDNRKKVQQQLISDIT 456  
DB

RESULT 7  
US-08-342-930-2  
Sequence 2, Application US/08342930  
Patent No. 5821084  
GENERAL INFORMATION:  
APPLICANT: OLIMSTED, ELIZABETH A.  
APPLICANT: MAURO, LAURA J.  
APPLICANT: DAVIS, ALAN R.  
APPLICANT: DIXON, JACK E.  
TITLE OF INVENTION: OSTEOBLAST-TESTICULAR PROTEIN TYROSINE  
NUMBER OF SEQUENCES: 13  
CORRESPONDENCE ADDRESS:  
ADDRESSEE: MORRISON & FORSTER  
STREET: 755 Page Mill Road  
City: Palo Alto  
STATE: CA  
COUNTRY: USA  
ZIP: 94304-1018  
COMPUTER READABLE FORM:  
MEDIUM TYPE: Floppy disk  
COMPUTER: IBM PC compatible  
OPERATING SYSTEM: PC-DOS/MS-DOS  
SOFTWARE: Patentin Release #1.0, Version #1.30  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/08/342,930  
FILING DATE: 21-NOV-1994  
CLASSIFICATION: 435  
ATTORNEY/AGENT INFORMATION:  
NAME: KOSKI, ANTOINETTE F.  
REGISTRATION NUMBER: 34,202  
REFERENCE/DOCKET NUMBER: 20344-20975.00  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: (415) 813-5600  
TELEFAX: (415) 494-0792  
TELEX: 706141  
INFORMATION FOR SEQ ID NO: 2:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 1711 amino acids  
TYPE: amino acid  
TOPOLOGY: linear  
MOLECULE TYPE: protein  
US-08-342-930-2

Query Match 6.7%; Score 80.5; DB 2; Length 1711;  
Best Local Similarity 29.6%; Pred. No. 19;

Matches 21; Conservative 8; Mismatches 17; Indels 25; Gaps 1;  
QY 36 SMENDTVYKSSPPLGAGLGAEPAGPQLPSWLQPRCAVFOCAQHVLADSVHLAW 95  
Db 753 SMOGSAVYVNLAMPSPGLGQGA-----CHQQLSDAGHLSW 787  
QY 96 DLSSSLAVVF 106  
Db 788 EQPLKLGQELF 798

## RESULT 8

US-09-018-576-3  
; Sequence 3, Application US/09018576  
; Patent No. 5968800  
; GENERAL INFORMATION:  
; APPLICANT: Gerhold, David L.  
; TITLE OF INVENTION: CYCLIN-DEPENDENT PROTEIN KINASE  
; NUMBER OF SEQUENCES: 14  
; CORRESPONDENCE ADDRESS:  
; ADDRESSEE: Merck & Co., Inc.  
; STREET: P.O. Box 2000, RY60-30  
; CITY: Rahway  
; STATE: NJ  
; COUNTRY: US  
; ZIP: 07065-0907  
; COMPUTER READABLE FORM:  
; MEDIUM TYPE: Floppy disk  
; COMPUTER: IBM PC compatible  
; OPERATING SYSTEM: PC-DOS/MS-DOS  
; SOFTWARE: Patentin Release #1.0, Version #1.30  
; CURRENT APPLICATION DATA:  
; APPLICATION NUMBER: US/09/018,576  
; FILING DATE:  
; CLASSIFICATION:  
; ATTORNEY/AGENT INFORMATION:  
; NAME: Hand, J. Mark  
; REGISTRATION NUMBER: 36,545  
; REFERENCE/DOCKET NUMBER: 19885Y  
; TELECOMMUNICATION INFORMATION:  
; TELEPHONE: 732/594-3905  
; TELEFAX: 732/594-4720  
; INFORMATION FOR SEQ. ID NO: 3:  
; SEQUENCE CHARACTERISTICS:  
; LENGTH: 325 amino acids  
; TYPE: amino acid  
; STRANDEDNESS: single  
; TOPOLOGY: linear  
; MOLECULE TYPE: protein  
; US-09-018-576-3

Query Match 6.6%; Score 78.5; DB 2; Length 325;  
Best Local Similarity 23.2%; Pred. No. 2.5;  
Matches 56; Conservative 24; Mismatches 78; Indels 83; Gaps 12;

QY 22 GGERAIQASTTSMENDTVYKSSPPLGAGLG-----AEEPAAG 63  
Db 43 GPPNQLKEIRKALQEME-DNQYVOLKAVFPFGGGFVLAFEFMLSDLAEVVHQAQRPPLAQ 101  
QY 64 POLPSWLOPERCAVFOCAQHVLADSVHLAMDLSRSLGAVVFSRYTNVV-----LEAPF 119  
Db 102 AQVKSXYLQMLKGV---AFCHA-----NNIVHRNLKPANL 133  
QY 120 LVGIEGSLKSGSTYMLPFGSGGIPVGFHLYSTHAALALRGHFCLSND----- 167  
Db 134 LISASGQKIKIDFGLARVFS---PDGSRLY-THQ--VATRSVGCIMGELNLSPLFPQKN 187  
QY 168 --KWCYLLKTKKALVNASSEM-DIQNVPLSEKIAELKEKIVLTHNRLLKSIKMLISEVTPDQ 224  
Db 188 DIEQLCYVLRILGTPNPQVWPELTLPDYNKIS-FKEQVPM-----LEEVLPDV 236  
QY 225 S 225

Db 237 S 237

## RESULT 9

US-09-018-576-12  
; Sequence 12, Application US/09018576  
; Patent No. 5968800  
; GENERAL INFORMATION:  
; APPLICANT: Gerhold, David L.  
; TITLE OF INVENTION: CYCLIN-DEPENDENT PROTEIN KINASE  
; NUMBER OF SEQUENCES: 14  
; CORRESPONDENCE ADDRESS:  
; ADDRESSEE: Merck & Co., Inc.  
; STREET: P.O. Box 2000, RY60-30  
; CITY: Rahway  
; STATE: NJ  
; COUNTRY: US  
; ZIP: 07065-0907  
; COMPUTER READABLE FORM:  
; MEDIUM TYPE: Floppy disk  
; COMPUTER: IBM PC compatible  
; OPERATING SYSTEM: PC-DOS/MS-DOS  
; SOFTWARE: Patentin Release #1.0, Version #1.30  
; CURRENT APPLICATION DATA:  
; APPLICATION NUMBER: US/09/018,576  
; FILING DATE:  
; CLASSIFICATION:  
; ATTORNEY/AGENT INFORMATION:  
; NAME: Hand, J. Mark  
; REGISTRATION NUMBER: 36,545  
; REFERENCE/DOCKET NUMBER: 19885Y  
; TELECOMMUNICATION INFORMATION:  
; TELEPHONE: 732/594-3905  
; TELEFAX: 732/594-4720  
; INFORMATION FOR SEQ. ID NO: 12:  
; SEQUENCE CHARACTERISTICS:  
; LENGTH: 325 amino acids  
; TYPE: amino acid  
; STRANDEDNESS: single  
; TOPOLOGY: linear  
; MOLECULE TYPE: protein  
; US-09-018-576-12

Query Match 6.6%; Score 78.5; DB 2; Length 325;  
Best Local Similarity 23.2%; Pred. No. 2.5;  
Matches 56; Conservative 24; Mismatches 78; Indels 83; Gaps 12;

QY 22 GGERAIQASTTSMENDTVYKSSPPLGAGLG-----AEEPAAG 63  
Db 43 GPPNQLKEIRKALQEME-DNQYVOLKAVFPFGGGFVLAFEFMLSDLAEVVHQAQRPPLAQ 101  
QY 64 POLPSWLOPERCAVFOCAQHVLADSVHLAMDLSRSLGAVVFSRYTNVV-----LEAPF 119  
Db 102 AQVKSXYLQMLKGV---AFCHA-----NNIVHRNLKPANL 133  
QY 120 LVGIEGSLKSGSTYMLPFGSGGIPVGFHLYSTHAALALRGHFCLSND----- 167  
Db 134 LISASGQKIKIDFGLARVFS---PDGSRLY-THQ--VATRSVGCIMGELNLSPLFPQKN 187  
QY 168 --KWCYLLKTKKALVNASSEM-DIQNVPLSEKIAELKEKIVLTHNRLLKSIKMLISEVTPDQ 224  
Db 188 DIEQLCYVLRILGTPNPQVWPELTLPDYNKIS-FKEQVPM-----LEEVLPDV 236  
QY 225 S 225  
Db 237 S 237

RESULT 10  
US-09-248-137-3  
; Sequence 3, Application US/09248137  
; Patent No. 6030788  
; GENERAL INFORMATION:

APPLICANT: Gerhold, David L.  
TITLE OF INVENTION: CYCLIN-DEPENDENT PROTEIN KINASE  
NUMBER OF SEQUENCES: 14  
CORRESPONDENCE ADDRESS:  
ADDRESSEE: Merck & Co., Inc.  
STREET: P.O. Box 2000, RY60-30  
CITY: Rahway  
STATE: NJ  
COUNTRY: US  
ZIP: 07065-0907  
COMPUTER READABLE FORM:  
MEDIUM TYPE: Floppy disk  
COMPUTER: IBM PC compatible  
OPERATING SYSTEM: PC-DOS/MS-DOS  
SOFTWARE: Patentin Release #1.0, Version #1.30  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/09/248,137  
FILING DATE:  
CLASSIFICATION:  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: 09/018,576  
FILING DATE:  
ATTORNEY/AGENT INFORMATION:  
NAME: Hand, J. Mark  
REGISTRATION NUMBER: 36,545  
REFERENCE/DOCKET NUMBER: 19885Y  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: 732/594-3905  
TELEFAX: 732/594-4720  
INFORMATION FOR SEQ ID NO: 3:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 325 amino acids  
TYPE: amino acid  
STRANDEDNESS: single  
TOPOLOGY: linear  
MOLECULE TYPE: protein  
US-09-248-137-3

Query Match 6.6%; Score 78.5; DB 3; Length 325;  
Best Local Similarity 23.2%; Pred. No. 2.5;  
Matches 56; Conservative 24; Mismatches 78; Indels 83; Gaps 12;

QY 22 GGERALIDQASTTSMEMTDYVYKSSPIGPAGLG-----AEEPAG 63  
DB 43 GPNQALREIKALOEME-DNQYVVLKAVFPFGGGLVLAPEFMLSGLAEVYRAQRLAQ 101  
QY 64 POLPSWLOPERCAVFCQACQCHAVLADSVHLAMDLSRSLGAVFSRVTNVY---LEAPF 119  
DB 102 AQVKSYIQMLKGV---AFCHA-----NNIVHRDLKPNAL 133  
QY 120 LVGIESGLKSTYNLLFCSCGGLPVGFHLVSTHAALALRGHPCLSSD----- 167  
DB 134 LISASGQLKADFGIARVFS---PDGSRILY-THQ--VATRSVGCIMGELLNGSPLFPKGN 187  
QY 168 --KWCYLLKTKRAIVNASEM-DIQNVPLSEKIALEKIKVLTNRRLKSLMKILSEVTPDQ 224  
DB 188 DIFQLCYVILGLTPNPQVWPELTLPDYNKIS-FKQVPM-----LEEVLPDV 236

QY 225 S 225  
DB 237 S 237

RESULT 11  
US-09-248-137-12  
Sequence 12, Application US/09248137  
Patent No. 6030788  
GENERAL INFORMATION:  
APPLICANT: Gerhold, David L.  
TITLE OF INVENTION: CYCLIN-DEPENDENT PROTEIN KINASE  
NUMBER OF SEQUENCES: 14  
CORRESPONDENCE ADDRESS:  
ADDRESSEE: Merck & Co., Inc.

STREET: P.O. Box 2000, RY60-30  
CITY: Rahway  
STATE: NJ  
COUNTRY: US  
ZIP: 07065-0907  
COMPUTER READABLE FORM:  
MEDIUM TYPE: Floppy disk  
COMPUTER: IBM PC compatible  
OPERATING SYSTEM: PC-DOS/MS-DOS  
SOFTWARE: Patentin Release #1.0, Version #1.30  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/09/248,137  
FILING DATE:  
CLASSIFICATION:  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: 09/018,576  
FILING DATE:  
ATTORNEY/AGENT INFORMATION:  
NAME: Hand, J. Mark  
REGISTRATION NUMBER: 36,545  
REFERENCE/DOCKET NUMBER: 19885Y  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: 732/594-3905  
TELEFAX: 732/594-4720  
INFORMATION FOR SEQ ID NO: 12:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 325 amino acids  
TYPE: amino acid  
STRANDEDNESS: single  
TOPOLOGY: linear  
MOLECULE TYPE: protein  
US-09-248-137-12

Query Match 6.6%; Score 78.5; DB 3; Length 325;  
Best Local Similarity 23.2%; Pred. No. 2.5;  
Matches 56; Conservative 24; Mismatches 78; Indels 83; Gaps 12;

QY 22 GGERALIDQASTTSMEMTDYVYKSSPIGPAGLG-----AEEPAG 63  
DB 43 GPNQALREIKALOEME-DNQYVVLKAVFPFGGGLVLAPEFMLSGLAEVYRAQRLAQ 101  
QY 64 POLPSWLOPERCAVFCQACQCHAVLADSVHLAMDLSRSLGAVFSRVTNVY---LEAPF 119  
DB 102 AQVKSYIQMLKGV---AFCHA-----NNIVHRDLKPNAL 133  
QY 120 LVGIESGLKSTYNLLFCSCGGLPVGFHLVSTHAALALRGHPCLSSD----- 167  
DB 134 LISASGQLKADFGIARVFS---PDGSRILY-THQ--VATRSVGCIMGELLNGSPLFPKGN 187  
QY 168 --KWCYLLKTKRAIVNASEM-DIQNVPLSEKIALEKIKVLTNRRLKSLMKILSEVTPDQ 224  
DB 188 DIFQLCYVILGLTPNPQVWPELTLPDYNKIS-FKQVPM-----LEEVLPDV 236

QY 225 S 225  
DB 237 S 237

RESULT 12  
US-09-252-991A-19049  
Sequence 19049, Application US/09252991A  
Patent No. 6551795  
GENERAL INFORMATION:  
APPLICANT: Marc J. Rubenfield et al.  
TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO PSEUDOMONAS  
TITLE OF INVENTION: AERGINOSA FOR DIAGNOSTICS AND THERAPEUTICS  
FILE REFERENCE: 107196,136  
CURRENT APPLICATION NUMBER: US/09/252,991A  
PRIOR FILING DATE: 1999-02-18  
CURRENT FILING DATE: 1998-02-18  
PRIOR APPLICATION NUMBER: US 60/074,788  
PRIOR FILING DATE: 1998-07-27

NUMBER OF SEQ ID NOS: 33142  
SEQ ID NO 19049  
LENGTH: 190  
TYPE: PRT  
ORGANISM: Pseudomonas aeruginosa  
US-09-252-991A-19049

Query Match  
Best Local Similarity 6.5%; Score 78; DB 4; Length 190;  
Matches 24; Conservative 6; Mismatches 29; Indels 14; Gaps 3;

QY 4 QPLRRSRCAFPFGDPCGGTERAIDQASFTTSMWDTQVYKSSPLGPAGLGAEEPAAG 63  
DB 94 RPHCTRRCVAPPAAGYFG--RAGDRAS-----VARSTGGLPAPAPGRPRGHC- 139  
QY 64 POLPSWLOPERCA 76  
DB 140 PAPRPMRPRRCA 152

## RESULT 13

US-09-537-682-1  
Sequence 1, Application US/09537682  
Patent No. 6303357  
GENERAL INFORMATION:  
APPLICANT: TAKEUCHI, Kenichi  
APPLICANT: KOIDE, Yoshinao  
APPLICANT: NAKANISHI, Yuji  
APPLICANT: SUZUKI, Satoru  
TITLE OF INVENTION: L-ALPHA-GLYCEROPHOSPHATE OXIDASE GENE, RECOMBINANT  
TITLE OF INVENTION: DNA, AND METHOD FOR PRODUCING MODIFIED  
FILE REFERENCE: A20-121814C/KI  
CURRENT APPLICATION NUMBER: US/09/537,682  
CURRENT FILING DATE: 2000-03-29  
NUMBER OF SEQ ID NOS: 2  
SOFTWARE: PatentIn Ver. 2.1  
SEQ ID NO 1  
LENGTH: 607  
TYPE: PRT  
ORGANISM: Enterococcus faecium No. 6303357 7044  
US-09-537-682-1

Query Match  
Best Local Similarity 6.4%; Score 77; DB 4; Length 607;  
Matches 49; Conservative 30; Mismatches 76; Indels 56; Gaps 10;

QY 19 DFCGGTERAIDQASFTTSM-----EMDTQVYKSSPLGP---AGLGAEEPAAGQ 65  
DB 51 DPAEGT-----SSRSTKLVHGGIRYLKTFDVEVV--ADTVGERAVVQGIAPHIPKDDM 102  
QY 66 LPSWLOPERCAVFOCAQCHAVLADSVHLAMDLSRSLGAVVFSRVN-----NVLEAPF 119  
DB 103 LPIYEDRGATTFNMF-----SVKVMOLYDLANTVGTKNYITLTPREVLREPF 154  
QY 120 LVIGESLKGSTVYNNLFCGSGCIPIYGF-----HLYSTHAALALRGHCLSSDKMVCYL 174  
DB 155 LK--KEGLKA-----GVYLDFFRNDARLVTDINIKKAABDGAVLVSMAKAVGFLY 202  
QY 175 KTKAIVNASEMDIONVPLSEKIAELKEKIVL 205  
DB 203 EGDQIVGVKARDL-----LTDEVIETKSKLVI 229

## RESULT 14

US-08-520-933-3  
Sequence 3, Application US/08520933  
Patent No. 5981194  
GENERAL INFORMATION:  
APPLICANT: Jefferies, Wilfred A.  
APPLICANT: McGeer, Patrick L.  
APPLICANT: Rothenberg, Sylvia  
APPLICANT: Food, Michael R.

APPLICANT: Yamada, Tatsuo  
APPLICANT: Kennard, Malcolm  
TITLE OF INVENTION: Use of p97 and Iron Binding Proteins  
TITLE OF INVENTION: as Diagnostic and Therapeutic Agents  
NUMBER OF SEQUENCES: 11  
CORRESPONDENCE ADDRESS:  
ADDRESSER: Bereskin & Parr  
STREET: 40 King Street West  
CITY: Toronto  
STATE: Ontario  
COUNTRY: Canada  
ZIP: M5H 3Y2  
COMPUTER READABLE FORM:  
MEDIUM TYPE: Floppy disk  
COMPUTER: IBM PC compatible  
OPERATING SYSTEM: PC-DOS/MS-DOS  
SOFTWARE: PatentIn Release #1.0, Version #1.25  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/08/520,933  
FILING DATE: August 31, 1995  
CLASSIFICATION: 424  
ATTORNEY/AGENT INFORMATION:  
NAME: Shona S. ModlarMid  
REGISTRATION NUMBER: 38,798  
REFERENCE/DOCKET NUMBER: 7685-006  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: 416-364-7311  
TELEFAX: 416-361-1398  
TELEX: 06-23115  
INFORMATION FOR SEQ ID NO: 3:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 719 amino acids  
TYPE: amino acid  
TOPOLOGY: linear  
MOLECULE TYPE: protein  
US-08-520-933-3

Query Match  
Best Local Similarity 6.4%; Score 77; DB 2; Length 719;  
Matches 59; Conservative 25; Mismatches 87; Indels 78; Gaps 15;

QY 17 RGPFCGTERAIDQASFTTSMK-----DTQVYKSSPLGPAGLGAEEPAAGP 64  
DB 171 RGDSSG--EGVCDKSPLERYDYSGAFRCIARAGAVARVSKSTVL-----EXTDCK 220  
QY 65 QLPWLOP-----ERCAVFOCAQCH--AVLADSVHLAMDLSRSLGAVVFSRV 109  
DB 221 TLPWGOALLSQDFELCRDGRADVTWRQCHLARVPAAVAVRADTD---GGLIF-RL 276  
QY 110 TNNVYLEAPFLVGTIGS-----LKGSTY--NLLFCGSGG--IPVGFHYST--HAALAA 157  
DB 277 LN-----EQRLFSSHSGSSFQWSSSEAYGQKDLFFDSTSELVPIATQYEAHLGHHYLA 332  
QY 158 LRGHFCLSSDKMVCYL-----LTKAIVNASEMDIONVPLSEKIAELKEKIVLTHNLKSL 213  
DB 333 MKGLIC-DPNRLPPLRWCVLSTPEIQCGDM-----AVAFRRQRLKEP 375  
QY 214 MKILSEVTP 222  
DB 376 IQCVSAKSP 384

## RESULT 15

US-09-285-040-3  
Sequence 3, Application US/09285040  
Patent No. 6455494  
GENERAL INFORMATION:  
APPLICANT: Jefferies, Wilfred A.  
APPLICANT: McGeer, Patrick L.  
APPLICANT: Rothenberg, Sylvia  
APPLICANT: Food, Michael R.  
APPLICANT: Yamada, Tatsuo  
APPLICANT: Kennard, Malcolm

TITLE OF INVENTION: Use of p97 and Iron Binding Proteins  
TITLE OF INVENTION: as Diagnostic and Therapeutic Agents  
NUMBER OF SEQUENCES: 11  
CORRESPONDENCE ADDRESS:  
ADDRESSEE: Bereskin & Parr  
STREET: 40 King Street West  
CITY: Toronto  
STATE: Ontario  
COUNTRY: Canada  
ZIP: M5H 3Y2  
COMPUTER READABLE FORM:  
MEDIUM TYPE: Floppy disk  
COMPUTER: IBM PC compatible  
OPERATING SYSTEM: PC-DOS/MS-DOS  
SOFTWARE: Patent Release #1.0, Version #1.25  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/09/285,040  
FILING DATE: 01-APR-1999  
CLASSIFICATION: 435  
ATTORNEY/AGENT INFORMATION:  
NAME: Gravelle, Micheline  
REGISTRATION NUMBER: 40,261  
REFERENCE/DOCKET NUMBER: 7685-032  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: 416-364-7311  
TELEFAX: 416-361-1398  
INFORMATION FOR SEQ ID NO: 3:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 719 amino acids  
TYPE: amino acid  
TOPOLOGY: linear  
MOLECULE TYPE: protein  
US-09-285-040-3

Query Match 6.4%; Score 77; DB 4; Length 719;  
Best Local Similarity 23.7%; Pred. No. 12;  
Matches 59; Conservative 25; Mismatches 87; Indels 78; Gaps 15;

QY 17 RDGFCGTEPAIDQASFTSMEM-----DTQVKGSSPIGPAAGAEPPAAGP 64  
DB 171 RDSSSG--EGVCDKSLERYDYISGAFLRCLAGAGDVAFVKSITVL-----ENTDCK 220  
QY 65 QLPFWLQP-----ERCAVFCQAOCH--AVIADSVHLAWDISRLGAVVFSRV 109  
DB 221 TLPSMGQALISQDFELLCRDGSRADYTEMROCHLAPPAHVVVRADTD--GLIF-RL 276  
QY 110 TNNVVLAEAPFLVIGES---LKGSTY--NLLFCGSCG--IPVGFHYST--HAALAA 157  
DB 277 LN---EGORLFSHSGSSFQMFSSSEAYGQKDLFFKDTSELVPIATQTYEAMLGHEYLHA 332  
QY 158 LRGHFLCLSDDKWVCYL---LKTKAIVNASENDIQVPISEKIAELKEKIVLTHNRKSL 213  
DB 333 MKGLLC-DPNRLPPYLRWCYVLTPEIQKGDG-----AAFRQRRLKPE 375  
QY 214 MKLISEVTP 222  
DB 376 IOCVSAKSP 384

Search completed: July 20, 2004, 11:10:45  
Job time : 21 secs



GenCore version 5.1.6  
Copyright (c) 1993 - 2004 Compugen Ltd.

OM protein - protein search, using sw model

Run on: July 20, 2004, 11:09:23 ; Search time 46 Seconds  
(without alignments)  
1556.009 Million cell updates/sec

Title: US-09-942-052A-728

Perfect score: 1198  
Sequence: 1 MAAPLHRRCATPPRGDF.....IKSLMKILSEVTPDQSKPEN 229

Scoring table:  
BLOSUM62  
Gapop 10.0 , Gapext 0.5

Searched: 1285356 seqs, 312560742 residues

Total number of hits satisfying chosen parameters: 1285356

Minimum DB seq length: 0  
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%  
Maximum Match 100%  
Listing first 45 summaries

Database :

Published Applications\_AA:\*

- 1: /cgn2\_6/ptodata/1/pubpaa/US07\_PUBCOMB.pep:\*
- 2: /cgn2\_6/ptodata/1/pubpaa/PCT\_NEW\_PUB.pep:\*
- 3: /cgn2\_6/ptodata/1/pubpaa/US06\_NEW\_PUB.pep:\*
- 4: /cgn2\_6/ptodata/1/pubpaa/US06\_PUBCOMB.pep:\*
- 5: /cgn2\_6/ptodata/1/pubpaa/US07\_NEW\_PUB.pep:\*
- 6: /cgn2\_6/ptodata/1/pubpaa/PCT\_US\_PUBCOMB.pep:\*
- 7: /cgn2\_6/ptodata/1/pubpaa/US08\_NEW\_PUB.pep:\*
- 8: /cgn2\_6/ptodata/1/pubpaa/US08\_PUBCOMB.pep:\*
- 9: /cgn2\_6/ptodata/1/pubpaa/US09\_PUBCOMB.pep:\*
- 10: /cgn2\_6/ptodata/1/pubpaa/US09\_PUBCOMB.pep:\*
- 11: /cgn2\_6/ptodata/1/pubpaa/US09\_NEW\_PUB.pep:\*
- 12: /cgn2\_6/ptodata/1/pubpaa/US10\_PUBCOMB.pep:\*
- 13: /cgn2\_6/ptodata/1/pubpaa/US10\_PUBCOMB.pep:\*
- 14: /cgn2\_6/ptodata/1/pubpaa/US10\_PUBCOMB.pep:\*
- 15: /cgn2\_6/ptodata/1/pubpaa/US10\_NEW\_PUB.pep:\*
- 16: /cgn2\_6/ptodata/1/pubpaa/US60\_NEW\_PUB.pep:\*
- 17: /cgn2\_6/ptodata/1/pubpaa/US60\_PUBCOMB.pep:\*
- 18: /cgn2\_6/ptodata/1/pubpaa/US60\_PUBCOMB.pep:\*

Prod. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

#### SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	1198	100.0	229	US-09-942-052-728	Sequence 728, App
2	1198	100.0	229	US-09-942-052-729	Sequence 729, App
3	1198	100.0	229	US-09-942-052-730	Sequence 730, App
4	1198	100.0	229	US-09-942-052-731	Sequence 731, App
5	138.5	11.6	164	US-09-942-052-707	Sequence 707, App
6	138	11.5	29	US-09-942-052-704	Sequence 704, App
7	113	9.4	233	US-09-890-688-82	Sequence 82, App
8	113	9.4	233	US-10-408-765A-969	Sequence 969, App
9	95.5	8.0	708	US-10-437-963-195427	Sequence 195427, App
10	92	7.7	361	US-10-437-963-154548	Sequence 154548, App
11	92	7.7	695	US-10-369-493-18389	Sequence 18389, App
12	89	7.4	664	US-09-780-525-2	Sequence 2, App
13	88.5	7.3	435	US-10-437-963-186569	Sequence 186569, App
14	87.5	7.3	848	US-10-437-963-155606	Sequence 155606, App
15	86.5	7.2	12	US-10-282-122A-51802	Sequence 51802, App

16	85	7.1	337	14	US-10-156-761-10954	Sequence 10954, A
17	84	7.0	225	12	US-10-424-599-223506	Sequence 223506, A
18	84	7.0	1902	16	US-10-437-963-139559	Sequence 139559, A
19	83.5	7.0	888	16	US-10-437-963-147897	Sequence 147897, A
20	83	6.9	663	12	US-10-282-122A-51673	Sequence 51673, A
21	82.5	6.9	278	12	US-10-425-114-42310	Sequence 42310, A
22	82.5	6.9	325	16	US-10-663-896-2	Sequence 2, App
23	82.5	6.9	502	12	US-10-424-599-226730	Sequence 226730, A
24	82.5	6.9	1043	15	US-10-369-493-9793	Sequence 9793, App
25	81.5	6.8	278	12	US-10-425-114-66220	Sequence 66220, A
26	81.5	6.8	457	13	US-10-195-158-6	Sequence 6, App
27	81.5	6.8	457	14	US-10-195-518-6	Sequence 6, App
28	81.5	6.8	1969	16	US-10-437-963-115342	Sequence 115342, A
29	81	6.8	859	12	US-10-210-281-60	Sequence 60, App
30	80	6.7	305	14	US-10-183-116-59	Sequence 59, App
31	80	6.7	544	12	US-10-424-599-158084	Sequence 158084, A
32	79.5	6.6	212	9	US-09-925-299-968	Sequence 968, App
33	79.5	6.6	212	10	US-09-925-299-968	Sequence 968, App
34	79	6.6	339	16	US-10-437-963-114922	Sequence 124922, A
35	79	6.6	482	14	US-10-160-764-81	Sequence 81, App
36	79	6.6	482	15	US-10-229-541A-87	Sequence 87, App
37	78.5	6.6	373	12	US-10-425-114-39137	Sequence 39137, A
38	78.5	6.6	452	9	US-09-771-161A-237	Sequence 237, App
39	78.5	6.6	452	12	US-10-403-161-8	Sequence 8, App
40	78.5	6.6	591	16	US-10-437-963-195425	Sequence 195425, A
41	78.5	6.6	855	16	US-10-437-963-148447	Sequence 148447, A
42	78.5	6.6	1259	16	US-10-437-963-102890	Sequence 102890, A
43	78.5	6.6	1563	12	US-10-282-122A-69748	Sequence 69748, A
44	78	6.5	162	14	US-10-017-161-1656	Sequence 1656, App
45	78	6.5	162	15	US-10-292-798-1320	Sequence 1320, App

#### ALIGNMENTS

RESULT 1  
US-09-942-052-728  
; Sequence 728, Application US/09942052  
; Publication No. US20030170626A1  
; GENERAL INFORMATION:  
; APPLICANT: Raitano, Arthur B.  
; APPLICANT: Paris, Mary  
; APPLICANT: Hubert, Rene S.  
; APPLICANT: Afari, Daniel  
; APPLICANT: Ge, Wangmao  
; APPLICANT: Chailita-Rid, Pia M.  
; TITLE OF INVENTION: NUCLEIC ACID AND CORRESPONDING PROTEIN ENTITLED 85PIB3  
; TITLE OF INVENTION: USEFUL IN TREATMENT AND DETECTION OF CANCER  
; FILE REFERENCE: 51158-20028.00  
; CURRENT APPLICATION NUMBER: US/09/942,052  
; CURRENT FILING DATE: 2001-08-28  
; PRIOR APPLICATION NUMBER: 60/228,432  
; PRIOR FILING DATE: 2000-08-28  
; NUMBER OF SEQ ID NOS: 744  
; SOFTWARE: Patent In Ver. 2.1  
; SEQ ID NO 728  
; LENGTH: 229  
; TYPE: PRT  
; ORGANISM: Unknown Organism  
; FEATURE:  
; OTHER INFORMATION: Description of Unknown Organism: 85PIB3/OIP5  
; OTHER INFORMATION: clone A protein  
US-09-942-052-728

Query Match: 100.0%; Score 1198; DB 10; Length 229;  
Best Local Similarity: 100.0%; Pred. No. 7.7e-117;  
Matches 229; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 MAAPLHRRCATPPRGDFCGGTERAIDASTTSMENOTVVKSSPGLGAEAP 60  
DB 1 MAAPLHRRCATPPRGDFCGGTERAIDASTTSMENOTVVKSSPGLGAEAP 60  
QY 61 AAGPOLPSWLOPRCAVFGCAQCAHVAIVLADSLAIVFVSRTNNVLEAPFL 120

```
Db 61 AAGPOLPSWLOPERCAVFOCAQCHAVLADSVHLAMDLSRSLGAVVFSRTNNVLEAPFL 120
QY 121 VGIEGSLKSTYNLLFCGSCGIPVGFHLYSTHAALALRGHFCSSDKMVCYLKTKAIY 180
Db 121 VGIEGSLKSTYNLLFCGSCGIPVGFHLYSTHAALALRGHFCSSDKMVCYLKTKAIY 180
QY 181 NASEMDIQNVPLSEKIAELKEKIVLTHNRKLSIMKILSEVTPDOSKPEP 229
Db 181 NASEMDIQNVPLSEKIAELKEKIVLTHNRKLSIMKILSEVTPDOSKPEP 229
```

## RESULT 2

```
US-09-942-052-729
; Sequence 729, Application US/09942052
; Publication No. US20030170626A1
; GENERAL INFORMATION:
; APPLICANT: Raitano, Arthur B.
; APPLICANT: Paris, Mary
; APPLICANT: Hubert, Rene S.
; APPLICANT: Afar, Daniel
; APPLICANT: Ge, Wangmao
; APPLICANT: Challita-Bid, Pia M.
; TITLE OF INVENTION: NUCLEIC ACID AND CORRESPONDING PROTEIN ENTITLED 85PIB3
; FILE REFERENCE: 51158-20028.00
; CURRENT APPLICATION NUMBER: US/09/942,052
; PRIOR FILING DATE: 2001-08-28
; PRIOR FILING DATE: 2000-08-28
; NUMBER OF SEQ ID NOS: 744
; SOFTWARE: Patentin Ver. 2.1
; SEQ ID NO 729
; LENGTH: 229
; TYPE: PRT
; ORGANISM: Unknown Organism
; FEATURE:
; OTHER INFORMATION: Description of Unknown Organism: 85PIB3/OIP5
; OTHER INFORMATION: protein sequence
US-09-942-052-729
```

```
Query Match 100.0%; Score 1198; DB 10; Length 229;
Best Local Similarity 100.0%; Pred. No. 7.7e-117;
Matches 229; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 MAOPLRHRSRCATPPRGDFCGGTERAIDQASFTTSMEMDTQVVGSSPLGPAIGAEPP 60
Db 1 MAOPLRHRSRCATPPRGDFCGGTERAIDQASFTTSMEMDTQVVGSSPLGPAIGAEPP 60
QY 61 AAGPOLPSWLOPERCAVFOCAQCHAVLADSVHLAMDLSRSLGAVVFSRTNNVLEAPFL 120
Db 61 AAGPOLPSWLOPERCAVFOCAQCHAVLADSVHLAMDLSRSLGAVVFSRTNNVLEAPFL 120
QY 121 VGIEGSLKSTYNLLFCGSCGIPVGFHLYSTHAALALRGHFCSSDKMVCYLKTKAIY 180
Db 121 VGIEGSLKSTYNLLFCGSCGIPVGFHLYSTHAALALRGHFCSSDKMVCYLKTKAIY 180
QY 181 NASEMDIQNVPLSEKIAELKEKIVLTHNRKLSIMKILSEVTPDOSKPEP 229
Db 181 NASEMDIQNVPLSEKIAELKEKIVLTHNRKLSIMKILSEVTPDOSKPEP 229
```

## RESULT 3

```
US-09-942-052-730
; Sequence 730, Application US/09942052
; Publication No. US20030170626A1
; GENERAL INFORMATION:
; APPLICANT: Raitano, Arthur B.
; APPLICANT: Paris, Mary
; APPLICANT: Hubert, Rene S.
; APPLICANT: Afar, Daniel
; APPLICANT: Ge, Wangmao
; APPLICANT: Challita-Bid, Pia M.
```

```
; TITLE OF INVENTION: NUCLEIC ACID AND CORRESPONDING PROTEIN ENTITLED 85PIB3
; FILE REFERENCE: 51158-20028.00
; CURRENT APPLICATION NUMBER: US/09/942,052
; PRIOR FILING DATE: 2001-08-28
; PRIOR FILING DATE: 2000-08-28
; NUMBER OF SEQ ID NOS: 744
; SOFTWARE: Patentin Ver. 2.1
; SEQ ID NO 730
; LENGTH: 229
; TYPE: PRT
; ORGANISM: Unknown Organism
; FEATURE:
; OTHER INFORMATION: Description of Unknown Organism: 85PIB3 protein
US-09-942-052-730
```

```
Query Match 100.0%; Score 1198; DB 10; Length 229;
Best Local Similarity 100.0%; Pred. No. 7.7e-117;
Matches 229; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 MAOPLRHRSRCATPPRGDFCGGTERAIDQASFTTSMEMDTQVVGSSPLGPAIGAEPP 60
Db 1 MAOPLRHRSRCATPPRGDFCGGTERAIDQASFTTSMEMDTQVVGSSPLGPAIGAEPP 60
QY 61 AAGPOLPSWLOPERCAVFOCAQCHAVLADSVHLAMDLSRSLGAVVFSRTNNVLEAPFL 120
Db 61 AAGPOLPSWLOPERCAVFOCAQCHAVLADSVHLAMDLSRSLGAVVFSRTNNVLEAPFL 120
QY 121 VGIEGSLKSTYNLLFCGSCGIPVGFHLYSTHAALALRGHFCSSDKMVCYLKTKAIY 180
Db 121 VGIEGSLKSTYNLLFCGSCGIPVGFHLYSTHAALALRGHFCSSDKMVCYLKTKAIY 180
QY 181 NASEMDIQNVPLSEKIAELKEKIVLTHNRKLSIMKILSEVTPDOSKPEP 229
Db 181 NASEMDIQNVPLSEKIAELKEKIVLTHNRKLSIMKILSEVTPDOSKPEP 229
```

## RESULT 4

```
US-09-942-052-731
; Sequence 731, Application US/09942052
; Publication No. US20030170626A1
; GENERAL INFORMATION:
; APPLICANT: Raitano, Arthur B.
; APPLICANT: Paris, Mary
; APPLICANT: Hubert, Rene S.
; APPLICANT: Afar, Daniel
; APPLICANT: Ge, Wangmao
; APPLICANT: Challita-Bid, Pia M.
; TITLE OF INVENTION: NUCLEIC ACID AND CORRESPONDING PROTEIN ENTITLED 85PIB3
; FILE REFERENCE: 51158-20028.00
; CURRENT APPLICATION NUMBER: US/09/942,052
; PRIOR FILING DATE: 2001-08-28
; PRIOR FILING DATE: 2000-08-28
; NUMBER OF SEQ ID NOS: 744
; SOFTWARE: Patentin Ver. 2.1
; SEQ ID NO 731
; LENGTH: 229
; TYPE: PRT
; ORGANISM: Homo sapiens
US-09-942-052-731
```

```
Query Match 100.0%; Score 1198; DB 10; Length 229;
Best Local Similarity 100.0%; Pred. No. 7.7e-117;
Matches 229; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 MAOPLRHRSRCATPPRGDFCGGTERAIDQASFTTSMEMDTQVVGSSPLGPAIGAEPP 60
Db 1 MAOPLRHRSRCATPPRGDFCGGTERAIDQASFTTSMEMDTQVVGSSPLGPAIGAEPP 60
```



```

QY 31 ASFTTSMEDTQVVGSSPLGAGAEPPAGPOLPSWLOPERCAVFOCAQHAVLADS 90
  |||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
Db 49 ASMSMSMEDASV---ADMERAOI--EEBAABE-----ERPLVFLCSCRRPLDSD 95
QY 91 VHLANDLSR-SIGAVVFSRVNNVNLVLEAPFLVIGISLKGSTYNNLLFCGSCGIPGFHLX 149
  |||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
Db 96 --LSWVASQEDTNCILRLCVCSCNVSVDKQKLSKREKENGCVLETLCCAGCSLNLGYVR 153
QY 150 STHAALALRGHFCISDKMVCYLL--KTKAIVNASEMDIONVPLSEKIAELKEKIVLTH 207
  |||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
Db 154 CTPKRLDYKRDPLFCISVAISYVIGSESEKOIV--SEDKELFNL---ESRVEIEKSLTQME 209
QY 208 NRLKSLMKILSE 219
  |||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
Db 210 DVLKALQMKLME 221
  |||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:

```

## RESULT 8

```

US-10-408-765A-969
; Sequence 969, Application US/10408765A
; Publication No. US20040101874A1
; GENERAL INFORMATION:
; APPLICANT: Ghosh, Soumitra S.
; APPLICANT: Pabry, Eoin D.
; APPLICANT: Zhang, Bing
; APPLICANT: Gibson, Bradford W.
; APPLICANT: Taylor, Steven W.
; APPLICANT: Glenn, Gary M.
; APPLICANT: Warnock, Dale E.
; TITLE OF INVENTION: TARGETS FOR THERAPEUTIC INTERVENTION
; TITLE OF INVENTION: IDENTIFIED IN THE MITOCHONDRIAL PROTEOME
; FILE REFERENCE: 660088.465
; CURRENT APPLICATION NUMBER: US/10/408,765A
; CURRENT FILING DATE: 2003-04-04
; NUMBER OF SEQ ID NOS: 3077
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 969
; LENGTH: 233
; TYPE: PRT
; ORGANISM: Homo sapiens
US-10-408-765A-969

```

```

Query Match
Best Local Similarity 9.4%; Score 113; DB 16; Length 233;
Matches 52; Conservative 30; Mismatches 88; Indels 22; Gaps 8;

```

```

QY 31 ASFTTSMEDTQVVGSSPLGAGAEPPAGPOLPSWLOPERCAVFOCAQHAVLADS 90
  |||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
Db 49 ASMSMSMEDASV---ADMERAOI--EEBAABE-----ERPLVFLCSCRRPLDSD 95
QY 91 VHLANDLSR-SIGAVVFSRVNNVNLVLEAPFLVIGISLKGSTYNNLLFCGSCGIPGFHLX 149
  |||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
Db 96 --LSWVASQEDTNCILRLCVCSCNVSVDKQKLSKREKENGCVLETLCCAGCSLNLGYVR 153
QY 150 STHAALALRGHFCISDKMVCYLL--KTKAIVNASEMDIONVPLSEKIAELKEKIVLTH 207
  |||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
Db 154 CTPKRLDYKRDPLFCISVAISYVIGSESEKOIV--SEDKELFNL---ESRVEIEKSLTQME 209
QY 208 NRLKSLMKILSE 219
  |||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
Db 210 DVLKALQMKLME 221
  |||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:

```

## RESULT 9

```

US-10-437-963-195427
; Sequence 195427, Application US/10437963
; Publication No. US20040123343A1
; GENERAL INFORMATION:
; APPLICANT: La Rosa, Thomas J.
; APPLICANT: Kovalic, David K.
; APPLICANT: Zhou, Yihua
; APPLICANT: Cao, Yongwei

```

```

; APPLICANT: Wu, Wei
; APPLICANT: Boukharov, Andrey A.
; APPLICANT: Barbazuk, Brad
; APPLICANT: Li, Ping
; TITLE OF INVENTION: Rice Nucleic Acid Molecules and Other Molecules Associated with
; TITLE OF INVENTION: Plants and Uses Thereof for Plant Improvement
; FILE REFERENCE: 38-21(53221)B
; CURRENT APPLICATION NUMBER: US/10/437,963
; CURRENT FILING DATE: 2003-05-14
; NUMBER OF SEQ ID NOS: 204966
; SEQ ID NO 195427
; LENGTH: 708
; TYPE: PRT
; ORGANISM: Oryza sativa
; FEATURE:
; OTHER INFORMATION: Clone ID: PAT_MRT4530_91377C.1.pep
US-10-437-963-195427

```

```

Query Match
Best Local Similarity 8.0%; Score 95.5; DB 16; Length 708;
Matches 71; Conservative 32; Mismatches 86; Indels 101; Gaps 17;

```

```

QY 6 LRRHRCATPPRGDF-----CGTERAIDQASFTTSMEW 39
  |||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
Db 225 LRRHRCATPPSASDASPPPPRQSIYITIGEKARAKARAKSGGTSASA-----SPTVST 279
QY 40 DTQVVGSSPLGAGAEPPAGPOLP---SM--LQPRCAVF-----QCAQCH 84
  |||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
Db 280 DVAVPVGSGQVWTSQ--PISDPAGPSLPRAVLTWELQVEMGRLLVAGANGIGREISEAR 338
QY 85 AYVADSVHLAMDLSRLG-----AVFSRVNNVLE-----APFLVIGISLKG 128
  |||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
Db 339 AETAAANALAERLVREIARAREDTYKRLVAGNERQSKLEDRLMSELGPNLSEIKSLR 358
QY 129 GSTYNNLL--FCGSCG---IPVQ---FHYSTHAALALRG-----HFCISDKM----- 169
  |||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
Db 399 -YTYTGLHQLAKEGVKSTIPVNLDEFSLTSSIAELATMGELPSKGTSLIAETSNGLY 457
QY 170 --VCYLLKTKAIVNNS--EMDIONV-----PLSEKIAELKEKIV 204
  |||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
Db 458 TGVCHVL---ACVRLSRPELDLREILDDQGAADTQREKVEVSDLSGSVL 504

```

## RESULT 10

```

US-10-437-963-154548
; Sequence 154548, Application US/10437963
; Publication No. US20040123343A1
; GENERAL INFORMATION:
; APPLICANT: La Rosa, Thomas J.
; APPLICANT: Kovalic, David K.
; APPLICANT: Zhou, Yihua
; APPLICANT: Cao, Yongwei
; APPLICANT: Wu, Wei
; APPLICANT: Boukharov, Andrey A.
; APPLICANT: Barbazuk, Brad
; APPLICANT: Li, Ping
; TITLE OF INVENTION: Rice Nucleic Acid Molecules and Other Molecules Associated with
; TITLE OF INVENTION: Plants and Uses Thereof for Plant Improvement
; FILE REFERENCE: 38-21(53221)B
; CURRENT APPLICATION NUMBER: US/10/437,963
; CURRENT FILING DATE: 2003-05-14
; NUMBER OF SEQ ID NOS: 204966
; SEQ ID NO 154548
; LENGTH: 361
; TYPE: PRT
; ORGANISM: Oryza sativa
; FEATURE:
; OTHER INFORMATION: Clone ID: PAT_MRT4530_54398C.1.pep
US-10-437-963-154548

```

```

Query Match
Best Local Similarity 7.7%; Score 92; DB 16; Length 361;
Matches 61; Conservative 28; Mismatches 92; Indels 60; Gaps 14;

```

```

QY 2 AAQPLRRHRCATPPRGDFCGTERAIDQASFTTSMEMDTQVYKSSPGLGPAGIGAEPA 61
DB 49 APTPLPRRAVRAAAADQSGSGTSSAPAVAST-----DYVVGSGSEAPPSG--PASPDP 102
QY 62 AGPOLP-----SW--LOPERCAVFOCAOCHAVLADSVHLAMDLSRLSGAVFSRYTNVVL 115
DB 103 AGRSPAAVLSMELQVEMGRLLEAGA--RVIGREIAEARGLERHM-----SELGNN--- 152
QY 116 EAPPLVIGESLKGSTYNLL--FCGSGCI-----PVGFHYSTHAALAL-----RGH 161
DB 153 -----LSEIRGSLR-VITYTGLHQLAGKCGIKSTIPAMPDEFSLTSLAEFLAAMEEIPSKH 207
QY 162 FCLSSDKM-----VCLTKTKAIVNASEMDIONV-----PLSEKIAELKEKI 203
DB 208 AARIGEENRIYIGACHITLACVRLAH--PELDLRILIDQGRASDARKDVEEVDLKGVS 266
QY 204 V 204
DB 267 L 267

RESULT 11
US-10-369-493-18389
; Sequence 18389, Application US/10369493
; Publication No. US20030233675A1
; GENERAL INFORMATION:
; APPLICANT: Cao, Yongwei
; APPLICANT: Hinkle, Gregory J.
; APPLICANT: Slater, Steven C.
; APPLICANT: Goldman, Barry S.
; APPLICANT: Chen, Xianfeng
; TITLE OF INVENTION: EXPRESSION OF MICROBIAL PROTEINS IN PLANTS FOR PRODUCTION OF
; TITLE OF INVENTION: PLANTS WITH IMPROVED PROPERTIES
; FILE REFERENCE: 38-10(52052)B
; CURRENT APPLICATION NUMBER: US/10/369,493
; CURRENT FILING DATE: 2003-02-28
; PRIOR APPLICATION NUMBER: US 60/360,039
; PRIOR FILING DATE: 2002-02-21
; NUMBER OF SEQ ID NOS: 47374
; SEQ ID NO 18389
; LENGTH: 695
; TYPE: PRP
; ORGANISM: Lactococcus lactis
US-10-369-493-18389

Query Match
Best local Similarity 7.7%; Score 92; DB 15; Length 695;
Matches 38; Conservative 28; Mismatches 50; Indels 34; Gaps 7;

QY 101 LGAVFSRYTNVNLAPPLVIGESLKGSTYNLLFCGSGCIPVGFHYSTHAALALRG 160
DB 101 LGTILP-----FYSGTFFPSGAKGELSKRKKPMMLITGILVAY-ANSVATYINSING 153
QY 161 HF-----CLSSDKVCCYLTKTKAIVNASE--MDI-----QNVPLSE-K 195
DB 154 HMGMTWTFELATITVIMLIGHLIMKALIMGADALDIASIVKKAHLKSGKVELSELK 213
QY 196 IAEIKERIVLTNRLKSLMKILSEVTPDQS 225
DB 214 VEDL--LVYENEKIPADGLILSEALVDES 241

RESULT 12
US-09-780-525-2
; Sequence 2, Application US/09780525
; Patent No. US20020004223A1
; GENERAL INFORMATION:
; APPLICANT: Bin-Bing Zhou
; APPLICANT: Yuan Zhu
; APPLICANT: Priya Chaturvedi
; APPLICANT: Mark R. Huile
; APPLICANT: Xiaotong Li

```

```

; TITLE OF INVENTION: PHAR1, A NEW RING FINGER PROTEIN
; FILE REFERENCE: GP-70668-C1
; CURRENT APPLICATION NUMBER: US/09/780,525
; CURRENT FILING DATE: 2001-02-09
; PRIOR APPLICATION NUMBER: 09/456,876
; PRIOR FILING DATE: 1999-12-08
; NUMBER OF SEQ ID NOS: 2
; SOFTWARE: FastSeq for Windows Version 3.0
; SEQ ID NO 2
; LENGTH: 664
; TYPE: PRP
; ORGANISM: HOMO SAPIENS
US-09-780-525-2

Query Match
Best local Similarity 7.4%; Score 89; DB 9; Length 664;
Matches 52; Conservative 26; Mismatches 74; Indels 74; Gaps 11;

QY 8 HRSRCATPPRGDFC-----GTERAIDQ-----SFTTSMEMDTQVYKSSP----- 50
DB 434 YRRQAAQPPH---CPAPEGEPAPQALGDAESTSVSLTTAVQDYVCPLOSHALCTCCFQ 490
QY 51 ---GPAGIGAEPPAGPOLPSMLQPERCAVFOCAOCHAVLADSVHLAMDLSR----- 99
DB 491 EMPRRAREODPRVAPQ-----QCAVC---LQPFCHLYGCTRTGCGYGCIA 534
QY 100 -----SIGAVFSRYTNVNLAPPLVIGESLKGSTY-NILFCGSGCIPVGFHYSTHA 153
DB 535 PFCELINLGDCKLDGVLYNNSESDILKNYLAT-RGLTWKNNL-----TES 578
QY 154 ALALRGHFCIS-----SDKAVCYLLTKTKAIVNASEMDIONVPLSE 194
DB 579 LVALRGVFLLDYKRVGTDTVLCCGGLSPRELTYQYRONIPASE 624

RESULT 13
US-10-437-963-186569
; Sequence 186569, Application US/10437963
; Publication No. US20040123343A1
; GENERAL INFORMATION:
; APPLICANT: La Rosa, Thomas J.
; APPLICANT: Kovalic, David K.
; APPLICANT: Zhou, Yinhua
; APPLICANT: Cao, Yongwei
; APPLICANT: Wu, Wei
; APPLICANT: Boukharov, Andrey A.
; APPLICANT: Barbazuk, Brad
; APPLICANT: Li, Ping
; TITLE OF INVENTION: Rice Nucleic Acid Molecules and Other Molecules Associated With
; TITLE OF INVENTION: Plants and Uses Thereof for Plant Improvement
; FILE REFERENCE: 38-21(53221)B
; CURRENT APPLICATION NUMBER: US/10/437,963
; CURRENT FILING DATE: 2003-05-14
; NUMBER OF SEQ ID NOS: 204966
; SEQ ID NO 186569
; LENGTH: 435
; TYPE: PRP
; ORGANISM: Oryza sativa
; FEATURE:
; OTHER INFORMATION: Clone ID: PAT_MRT4530_83357C.1.pep
US-10-437-963-186569

Query Match
Best local Similarity 7.4%; Score 88.5; DB 16; Length 435;
Matches 50; Conservative 27; Mismatches 82; Indels 59; Gaps 12;

QY 2 AAQPL--RH--RSRCATPPRGDFCGTERAIDQASFTTSMEMDTQVYKSSPGLGPAGIGA 57
DB 41 APEPLSCRHGRHRCADV-----GGAGRETERPPAPQ-----RESPSSGLDAL 88
QY 58 EEPDAGPOLPSWLOPERCAVFOCAOCHA--VLADSVHLAMDLSRLSGAVFSRYTNVVL 115
DB 89 EDSPQPGVPLILP-----LCRCYAKKICEYVVRTDVLVNH-----LNSVAIS 135

```

OY 116 EAPFLVIGESLKGSTYNILFCGSCGIPVGFHLYTHAALARGHFLCLSSDKWVCYLK 175  
Db 136 EGPFL-----SMRARRFLGSAS---ASVQKTEPF-----CATDSKRCY-LQ 173  
OY 176 TKALVNASMDION-----VPLSEKIAELKEKIVLTH 207  
Db 174 NGSFGITEDEGSSLYNFLYPSKELLPPDDKMSIFDH 211

## RESULT 14

US-10-437-963-155606  
; Sequence 155606, Application US/10437963  
; Publication No. US2004012343A1  
; GENERAL INFORMATION:  
; APPLICANT: La Rosa, Thomas J.  
; APPLICANT: Kovalic, David K.  
; APPLICANT: Zhou, Yihua  
; APPLICANT: Cao, Yongwei  
; APPLICANT: Wu, Wei  
; APPLICANT: Boukharov, Andrey A.  
; APPLICANT: Barbazov, Brad  
; APPLICANT: Li, Ping  
; TITLE OF INVENTION: Rice Nucleic Acid Molecules and Other Molecules Associated with  
; FILE REFERENCE: 38-21(53221)B  
; CURRENT APPLICATION NUMBER: US/10/437,963  
; CURRENT FILING DATE: 2003-05-14  
; NUMBER OF SEQ ID NOS: 204966  
; SEQ ID NO 155606  
; LENGTH: 848  
; TYPE: PRT  
; ORGANISM: *Oryza sativa*  
; FEATURE:  
; NAME/KEY: unsure  
; LOCATION: (1)..(848)  
; OTHER INFORMATION: unsure at all Xaa locations  
; FEATURE:  
; OTHER INFORMATION: Clone ID: PAT\_MRT4530\_55353C.1.pep  
; US-10-437-963-155606

Query Match 7.3%; Score 87.5; DB 16; Length 848;  
Best Local Similarity 22.1%; Pred. No. 9.8; Mismatches 109; Indels 97; Gaps 13;  
Matches 65; Conservative 23;  
OY 16 PRGDFCGTERAIDASFTTSMEMDTQVYKSSPLGAGLGAEP----- 60  
Db 85 PRGFLDGTGGVRRQTASCHLTRDLTDFLKSGCLGPRGRKRNPTPPPLGNGOEGRHL 144  
OY 61 AAGPOLPSMLQPERCAVPOCAQCH-----AVLADSVHAWDLR----- 99  
Db 145 TLGPVPVPGARERGRICLPQASGHDTPGFISPTSLVLRKTRIRIDEVAVNTIDEANEG 204  
OY 100 --SLGAVV--FSRVTVNNV-----VLEAPFLVIGESLKGSTYNILFCG----- 138  
Db 205 YVSGSVTEMSRQKAAAGVPAOSXPACXGIPVGVKGSV--LIFTARMGGLCHMTT 262  
OY 139 ---SCGIPV-----GFHLVSTH---AALAAARGHFCUSSDKWVCYL 173  
Db 263 GSKAMECGVKKVCIWGPWLAWDELGRBSGYOGLNHRPLVKAVIDGHLCEBDDAVNCLP 322  
OY 174 LKTRAIIVASMDIONVPLSEKIAELKEKIVLTHNRKLKMLKILSEVTPDOSKP 227  
Db 323 KLRTVSGSAS-----AKEAVKPAVQKEKIRS-VKVLSTVS-DLSLP 363

## RESULT 15

US-10-282-122a-51802  
; Sequence 51802, Application US/10282122A  
; Publication No. US20040029129A1  
; GENERAL INFORMATION:  
; APPLICANT: Wang, Liangsu  
; APPLICANT: Zamudio, Carlos

APPLICANT: Malone, Cheryl  
APPLICANT: Haselbeck, Robert  
APPLICANT: Ohlsen, Kari  
APPLICANT: Zvekind, Judith  
APPLICANT: Wall, Daniel  
APPLICANT: Trewick, John  
APPLICANT: Grant, Grant  
APPLICANT: Yamamoto, Robert  
APPLICANT: Forsyth, R.  
APPLICANT: Xu, H.  
; TITLE OF INVENTION: Identification of Essential Genes in Microorganisms  
; FILE REFERENCE: EPIRA.034A  
; CURRENT APPLICATION NUMBER: US/10/282,122A  
; CURRENT FILING DATE: 2003-02-20  
; PRIOR APPLICATION NUMBER: 60/191,078  
; PRIOR FILING DATE: 2000-03-21  
; PRIOR APPLICATION NUMBER: 60/206,848  
; PRIOR FILING DATE: 2000-05-23  
; PRIOR APPLICATION NUMBER: 60/207,727  
; PRIOR FILING DATE: 2000-05-26  
; PRIOR APPLICATION NUMBER: 60/230,335  
; PRIOR FILING DATE: 2000-09-06  
; PRIOR APPLICATION NUMBER: 60/230,347  
; PRIOR FILING DATE: 2000-09-09  
; PRIOR APPLICATION NUMBER: 60/242,578  
; PRIOR FILING DATE: 2000-10-23  
; PRIOR APPLICATION NUMBER: 60/253,625  
; PRIOR FILING DATE: 2000-11-27  
; PRIOR APPLICATION NUMBER: 60/257,931  
; PRIOR FILING DATE: 2000-12-22  
; PRIOR APPLICATION NUMBER: 60/267,636  
; PRIOR FILING DATE: 2001-02-09  
; PRIOR APPLICATION NUMBER: 60/269,308  
; PRIOR FILING DATE: 2001-02-16  
; Remaining Prior Application data removed - See File Wrapper or PALM.  
; NUMBER OF SEQ ID NOS: 78614  
; SOFTWARE: PatentIn version 3.1  
; SEQ ID NO 51802  
; LENGTH: 669  
; TYPE: PRT  
; ORGANISM: *Clostridium acetobutylicum*  
; US-10-282-122a-51802

Query Match 7.2%; Score 86.5; DB 12; Length 669;  
Best Local Similarity 27.4%; Pred. No. 9; Mismatches 45; Indels 31; Gaps 8;  
Matches 37; Conservative 22;  
OY 96 DLSR---SLGAVPERVNNVYLEAPFLVIGESLKGSTYNL-----LFCGSCGIPVGF 146  
Db 366 DGRKVKIGSRVFRNSNDVI---PEITGVTEETEGTNEIEAPTCIPYCGSEIVKEGV 422  
OY 147 HLSTHAALARGHFCUSSDKWVCYLKTRAIIVASMDIONVPLSEKIAELKEKIV 204  
Db 423 HL-----FC--ENTLSCKPQWKSIVHFAREAMNIGFSEKTAEOULFEK-- 465  
OY 205 LTHNRKLKMLKILSE 219  
Db 466 LNKISIDLYRITKE 480

Search completed: July 20, 2004, 11:15:23  
Job time : 47 secs

GenCore version 5.1.6  
Copyright (c) 1993 - 2004 Compugen Ltd.

OM protein - protein search, using sw model

Run on: July 20, 2004, 11:14:34 ; Search time 16 Seconds  
(without alignments)  
1376.741 Million cell updates/sec

Title: US-09-942-052A-728

Perfect score: 1198  
Sequence: 1 MAQPLHRRRCATPPRGDF.....LKSLMKILSEVTPDQSKREN 229

Scoring table: BIOSUM62  
Gapop 10.0 , Gapext 0.5

Searched: 283366 seqs, 96191526 residues

Total number of hits satisfying chosen parameters: 283366

Minimum DB seq length: 0  
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%  
Listing first 45 summaries

Database :  
1: pir1.\*  
2: pir2.\*  
3: pir3.\*  
4: pir4.\*

Pred. No. is the number of results predicted by chance to have a  
score greater than or equal to the score of the result being printed,  
and is derived by analysis of the total score distribution.

## SUMMARIES

Result No.	Score	% Match	Query Length	ID	Description
1	92	7.7	695	C86731	copper-potassium t
2	90.5	7.6	597	D71293	probable phosphori
3	89	7.4	306	T46399	hypothetical prote
4	88	7.3	250	A43623	kanamycin kinase (
5	86.5	7.2	669	A97229	NAD-dependent DNA
6	85	7.1	2194	J01977	glutamate synthase
7	84.5	7.1	261	AE2074	ferrichrome bindin
8	83	6.9	663	D97047	DNA ligase (NMD de
9	82.5	6.9	143	B69099	formate hydrogenly
10	82.5	6.9	493	A85433	sugar transporter
11	82	6.8	967	F87678	DNA polymerase I (
12	81.5	6.8	329	D90404	transpor protein,
13	81.5	6.8	457	C90919	glucuronide permea
14	81.5	6.8	457	B64918	glucuronide permea
15	81.5	6.8	457	C90919	glucuronide permea
16	81.5	6.8	457	H85767	glucuronide permea
17	81.5	6.8	640	T41977	hypothetical prote
18	81.5	6.8	640	T41977	hypothetical prote
19	81.5	6.8	640	T41977	hypothetical prote
20	80.5	6.7	1418	S40762	receptor kinase-11
21	80.5	6.7	1711	A55148	protein-tyrosine-p
22	80	6.7	469	S55167	IME2-dependent sig
23	79	6.6	386	H90789	probable aminometh
24	79	6.6	386	C85650	probable aminometh
25	78.5	6.6	437	C86823	GTP-binding protei
26	78.5	6.6	378	F83416	late competence op
27	78.5	6.6	578	F86484	probable hydroxyme
28	78	6.5	351	Q08EK3	Uth104 protein - hu
29	77.5	6.5	662	H82098	UDP-3-O-3-hydroxym
				T44036	hypothetical prote

30	77.5	6.5	1013	2	T31211	trwC protein homol
31	77.5	6.5	1643	2	T14274	versican precursor
32	77	6.4	381	2	AD1113	hypothetical prote
33	77	6.4	721	1	F87611	TonB-dependent rec
34	77	6.4	738	1	TFHDM	melanotransferrin
35	76.5	6.4	307	2	AG2017	glycerol-3-phospha
36	76.5	6.4	387	2	D69392	probable acyl-CoA
37	76.5	6.4	388	2	S57526	cellulase - Fibrob
38	76.5	6.4	764	2	AD3144	formate dehydrogen
39	76.5	6.4	764	2	H98143	cbbBc protein (U60
40	76.5	6.4	1554	2	T06370	probable DNA (cyto
41	76	6.3	367	2	AH0936	glycerol dehydroge
42	75.5	6.3	310	2	A55053	endothelial monocy
43	75.5	6.3	415	2	T46716	hypothetical prote
44	75.5	6.3	402	2	D70951	probable Urid - My
45	75.5	6.3	1534	2	S59604	DNA (cytosine-5-)-

## ALIGNMENTS

## RESULT 1

C86731  
copper-potassium transporting ATPase B copB [imported] - Lactococcus lactis subsp. lact  
C:Species: Lactococcus lactis subsp. lactis  
C:Date: 23-Mar-2001 #sequence\_revision 23-Mar-2001 #text\_change 03-Aug-2001  
C:Accession: C86731  
R:Bojoltin, A.; Winkler, P.; Manger, S.; Jallion, O.; Malarne, K.; Weissenbach, J.; Ehr  
Genome Res. 11, 731-753, 2001  
A:Title: The complete genome sequence of the lactic acid bacterium Lactococcus lactis s  
A:Reference number: A86625; MUID:21235186; PMID:11337471  
A:Accession: C86731  
A:Status: preliminary  
A:Molecule type: DNA  
A:Residues: 1-695 <STO>  
A:Cross-references: GB:AF005176; PID:q12723778; PIDN:AAK04949.1; GSPDB:GN00146  
A:Experimental source: strain IL1403  
C:Genetics:  
A:Gene: copB  
C:Superfamily: Enterococcus copper-transferring ATPase copB; ATPase nucleotide-binding

Query Match 7.7%; Score 92; DB 2; Length 695;

Best Local Similarity 25.3%; Pred No. 2.2; Mismatches 50; Indels 34; Gaps 7;

QY	101	IGAVFSRYTNVNVLEAPFLVIGESLKGSTYLLFCGSGGIPVGFHLSTHAAALRG 160
DB	101	LGTTIF-----FSGTPEFGAKGELKSRKPPAMMLITMGITVAY-AVSVAITMSLNG 153
QY	161	HF-----CLSDKAWCYLTKTKAYNASE--MDI-----GNVPESE-K 195
DB	154	HNGMNFWFELATLIVIMLIGHILIEMKATIGAGDAUKLASLVKKAHLKSGKVELSEUK 213
QY	196	IAELKEKIVLTNNRLKSLMKIILSEVTPDS 225
DB	214	VDPL--LIVKENKIPADGILISEALVDS 241

## RESULT 2

D71293  
probable phosphoribosylglycinamide formyltransferase - syphilis spirochete  
C:Species: Treponema pallidum subsp. pallidum (syphilis spirochete)  
C:Date: 24-Jul-1998 #sequence\_revision 24-Jul-1998 #text\_change 05-Nov-1999  
R:Fraser, C.M.; Norris, S.J.; Weinstock, G.M.; White, O.; Sutton, G.G.; Dodson, R.; Gw  
rson, J.; Khajak, H.; Richardson, D.; Howell, J.K.; Chidambaram, M.; Utterback, T.; Mc  
Sweeney, L.; Weidman, J.; Smith, H.O.; Venter, J.C.  
Science 281, 375-388, 1998  
A:Title: Complete genome sequence of Treponema pallidum, the syphilis spirochete.  
A:Reference number: A71250; MUID:98332770; PMID:9665876  
A:Accession: D71293  
A:Status: preliminary; nucleic acid sequence not shown; translation not shown  
A:Molecule type: DNA

A;Residues: 1-597 <COL>  
A;Cross-references: GB:AE001243; GB:AE000520; NID:g3322990; PIDN:AA065662.1; PID:g332299  
A;Experimental source: strain Nichols  
C;Genetics:  
A;Gene: TP0695

Query Match  
Best Local Similarity 7.6%; Score 90.5; DB 2; Length 597;  
Matches 51; Conservative 19; Mismatches 47; Indels 65; Gaps 12;

QY 12 CATPPRDPCGGTERAID-----QASFT-----TSNEMDTQVVKSGSSPLGPA 53  
DB 90 CALP--CHRLPATTNATDKTRMRACFTPARLRCPRTFLPDSFAMDT-----PPGHA 140  
QY 54 GLGAEPBAQPOLPSWLP-ERCAPQC--AOCNAVADSVHLAMDLSRLGAVFESRT 110  
DB 141 RLCSHLSHAGSLPPIVVKPTDMMGACGCTLAOCKDTLIMCAVARQPSRS----- 190  
QY 111 NNVVLEAPFLVIGESLKGSTYNLLFCGSCGIPVGFHLYSTHAALMRGHFCLSDRMV 170  
DB 191 GRVILR-EFIVGRFSLHG---LIFDGT-----LYVT--ALA-----DRHI 225  
QY 171 CY 172  
DB 226 CF 227

## RESULT 3

T46399

hypothetical protein DKFZp434N2420.1 - human (fragment)  
C;Species: Homo sapiens (man)  
C;Date: 04-Feb-2000 #sequence\_revision 04-Feb-2000 #text\_change 04-Feb-2000  
C;Accession: T46399

R;Oltenswelder, B.; Obermaier, B.; Mewes, H.W.; Gassenhuber, J.; Wiemann, S.  
submitted to the Protein Sequence Database, January 2000  
A;Reference number: Z23031  
A;Accession: T46399

A;Status: preliminary  
A;Molecule type: mRNA  
A;Residues: 1-306 <AAA>  
A;Cross-references: EMBL:AL137561  
A;Experimental source: adult testis; clone DKFZp434N2420  
C;Genetics:  
A;Note: DKFZp434N2420.1

Query Match  
Best Local Similarity 7.4%; Score 89; DB 2; Length 306;  
Matches 52; Conservative 26; Mismatches 74; Indels 74; Gaps 11;

QY 8 HRSRCATPPRGDFC-----GGERAIDQA-----SFTTSMEMDTQVVKSGSSPL----- 50  
DB 76 YRRQAQPPH---CPAPGEPGAPQALGAPSTSVSLTAVVDVCPQGSHALCTCCFQ 132  
QY 51 ---GPAGLGAEPBAQPOLPSWLP-ERCAPQCACCHAVLADSVHLAMDLSR----- 99  
DB 133 PMPDRARRBODPRVAPQ-----QCAVC---LQPFCHLYWGCTRTGCGYCLA 176  
QY 100 ----SLGAVFESRTNNVVLEAPFLVIGESLKGSTYNLLFCGSCGIPVGFHLYSTHA 153  
DB 177 PFCEINLGDCKLDGLVANNYSBSDLKNYLAT-RLITKKNM-----TES 220  
QY 154 ALAALRGHFCIS-----SDKRVCYILTKTAIVNASMDIQNVPLSE 194  
DB 221 LVALRGVFLISDYKVTGDTVLCCGCCGRSPRELTYQYRQNIIPASE 266

## RESULT 4

A43623

kanamycin kinase (EC 2.7.1.95) - Campylobacter jejuni  
C;Species: Campylobacter jejuni  
C;Date: 11-Feb-1993 #sequence\_revision 11-Feb-1993 #text\_change 08-Oct-1999  
C;Accession: A43623

R;Tenover, F.C.; Gilbert, T.; O'Hara, P.

Plasmid 22, 52-58, 1989  
A;Title: Nucleotide sequence of a novel kanamycin resistance gene,  $\alpha$ ph-7, from Campylobacter  
A;Reference number: A43623; MUID:89387451; PMID:2550983  
A;Accession: A43623

A;Status: preliminary  
A;Molecule type: DNA  
A;Residues: 1-250 <TEN>  
A;Cross-references: GB:M29953; GB:J03316; NID:g144186; PIDN:AAA76822.1; PID:g144187  
C;Superfamily: Kanamycin Kinase  
C;Keywords: phosphotransferase

Query Match  
Best Local Similarity 7.3%; Score 88; DB 2; Length 250;  
Matches 33; Conservative 18; Mismatches 46; Indels 18; Gaps 5;

QY 105 VFSRVNNVVLEAPFLVIGESLKGSTYNLLFCGSCGIPVGFHLYSTHAALMRGHFCLSDRMV 163  
DB 40 IFKITYSVRERAEEMMMWLSDKLKPEDV-----IYGVBRHSRYLTMSELGKHID 90  
QY 164 LSDKRVCYILTKTAIVNA-----SEMDIQNVPLSEKI-AELKKEIVLTHNRKSL 213  
DB 91 CFIDHPIXYI---ECLVNAHLQLAIDIRNCPSSKIDVRLKELKYLIDNRADI 142

## RESULT 5

A97229

NAD-dependent DNA ligase [imported] - Clostridium acetobutylicum  
C;Species: Clostridium acetobutylicum  
C;Date: 14-Sep-2001 #sequence\_revision 14-Sep-2001 #text\_change 30-Sep-2001  
C;Accession: A97229

R;Nolling, J.; Breton, G.; Omelchenko, M.V.; Markarova, K.S.; Zeng, Q.; Gibson, R.; Lee, J.; Daly, M.U.; Bennett, G.N.; Koonin, E.V.; Smith, D.R.  
J. Bacteriol. 183, 4823-4838, 2001  
A;Title: Genome Sequence and Comparative Analysis of the Solvent-Producing Bacterium Clostridium acetobutylicum  
A;Reference number: A96900; MUID:21359325; PMID:21359325  
A;Accession: A97229

A;Status: preliminary  
A;Molecule type: DNA  
A;Residues: 1-669 <KUR>  
A;Cross-references: GB:AE001437; PIDN:AAK80620.1; PID:g15025704; GSPDB:GN00168  
A;Experimental source: Clostridium acetobutylicum ATCC824  
C;Genetics:  
A;Gene: CAC2673  
C;Superfamily: polydeoxyribonucleotide synthase (NAD+)

Query Match  
Best Local Similarity 7.2%; Score 86.5; DB 2; Length 669;  
Matches 37; Conservative 22; Mismatches 45; Indels 31; Gaps 8;

QY 96 DLNR---SLGAVFESRTNNVVLEAPFLVIGESLKGSTYNLLFCGSCGIPVGF 146  
DB 366 DGRKKVKIGSRVFRNSNDVI---PEIMGVTEETEGTNEIEAPTICPYCGSEIVKGV 422  
QY 147 HLYSTHAALMRGHFCLSDRMVWVLYLTKTAIVNASMDIQNV-LEKTAI-LEKTI 204  
DB 423 HL-----FC--ENTLSCKPQVAKSVIHFARREANNIEGFEKTAIEQLFEK-- 465  
QY 205 LTHNRKSLMKTISE 219  
DB 466 LNIKSIDLYRITKE 480

## RESULT 6

J01977

glutamate synthase (NADH2) (EC 1.4.1.14) precursor [validated] - alfalfa  
C;Species: Medicago sativa (alfalfa)  
C;Date: 10-Sep-1999 #sequence\_revision 10-Sep-1999 #text\_change 03-Jun-2002  
C;Accession: J01977; PQ0551

R;Gregerson, R.G.; Miller, S.S.; Twary, S.N.; Gant, J.S.; Vance, C.P.  
Plant Cell 5, 215-226, 1993  
A;Title: Molecular characterization of NADH-dependent glutamate synthase from alfalfa  
A;Reference number: J01977; MUID:93200806; PMID:8453303  
A;Accession: J01977

A:Molecule type: mRNA  
 A:Residues: 1-2194 <GRE>  
 A:Cross-references: GB:L01660; NID:g166411; PIDN:AB46617.1; PID:g166412  
 A:Accession: P00551  
 A:Molecule type: protein  
 A:Residues: 102-114 <GR2>  
 C:Comment: This enzyme catalyzes the reductive transfer of the amido group of glutamine  
 C:Superfamily: glutamate synthase (NADH)  
 C:Keywords: 3Fe-4S; chloroplast; iron-sulfur protein; metalloprotein; NAD; oxidoreductase  
 F:1-101/Domain: propeptide #status predicted <PRO>  
 F:102-2194/Product: glutamate synthase (NADH) #status experimental <MAT>  
 F:102/Active site: Cys #status predicted  
 F:1246,1252,1257/Binding site: 3Fe-4S cluster (Cys) (covalent) #status predicted

Query Match 7.1%; Score 85; DB 1; Length 2194;  
 Best Local Similarity 24.7%; Pred. No. 42;

Matches 58; Conservative 22; Mismatches 79; Indels 76; Gaps 12;

QY 4 QPLRRSRCAAPPFGDPCGTERAIDQAS-----FTSMEMDTQVYKSSPLGPA 53  
 DB 1037 EPLADGSR--NPKRS-----AIKQVASGRFVSSYYLTNNADBLQIKMAQAKP----- 1082  
 QY 54 GLGAEP-----AAGPOLPSWLOPEKCAVFCQACQCHAVLADSVHLAMDLSR 99  
 DB 1083 GEGGELPEHKYIGDAITRNSTAGVGLIS--PPPHDIYS-----IEDLAQLTHDLKN 1133  
 QY 100 SLGAVFSRVNNVLEAPFLVIGESLKGSTYNLLFCG-----SCGIPVG 145  
 DB 1134 ANPA---ARISVKVSEAGVIVASGVYKGAHEHLISGHOGTGASRWTKGSKGLWE 1190  
 QY 146 PHLYSTHAALAA--LRGHFCLSSDKMVCYLLKT-----KATVNASEMDIQNVPL 192  
 DB 1191 LGIAETHQTIIVANDLRGRTLTQTDQ-----LKTGDVAIAALGAEYGFSTAPL 1241

# RESULT 7

ferrichrome binding protein of ABC transporter all12147 [imported] - Nostoc sp. (strain F AE2074  
 C:Species: Nostoc sp. PCC 7120  
 A>Note: Nostoc sp. strain PCC 7120 is a synonym of Anabaena sp. strain PCC 7120  
 C:Date: 14-Dec-2001 #sequence revision 14-Dec-2001 #text\_change 09-Dec-2002  
 C:Accession: AE2074  
 R:Kaneko, T.; Nakamura, Y.; Molk, C.P.; Kuritz, T.; Sasamoto, S.; Watanabe, A.; Iriguchi, Nakazaki, N.; Shimpo, S.; Sugimoto, M.; Takazawa, M.; Yamada, M.; Tabata, S  
 DNA Res. 8, 205-213, 2001  
 A:Title: Complete Genomic Sequence of the Filamentous Nitrogen-fixing Cyanobacterium Anabaena  
 A:Reference number: AB1807; MUID:21595285; PMID:11759840  
 A:Accession: AE2074  
 A:Status: preliminary  
 A:Molecule type: DNA  
 A:Residues: 1-361 <KUR>  
 A:Cross-references: GB:BA00019; PIDN:BA073846.1; PID:g17131238; GSPDB:GN00179  
 A:Experimental source: strain PCC 7120  
 C:Genetics:  
 A:Superfamily: ferrichrome-iron transport protein fecB

Query Match 7.1%; Score 84.5; DB 2; Length 361;  
 Best Local Similarity 26.1%; Pred. No. 4.8;

Matches 36; Conservative 17; Mismatches 36; Indels 49; Gaps 7;

QY 97 LSRSIGAVVF---SRVT---NNVLEAPFLVIGESLKGSTYNLLFCGSC-----GIPV 144  
 DB 81 ISHALGKVKIPKQKRVVLEENIILDSVALGVK-----PVGWVYQDDCENPRGID- 133  
 QY 145 GPHLYSTHAALALRGHFLSSDKMVCYLLKTAIVNASEMDIQNVPLSEKIAELKEKIV 204  
 DB 134 -----SD-----LLADVPGV-----NIGNPSTLEKILSLKPDIL 163  
 QY 205 LTHNRLKSLMKITSEVTP 222  
 DB 164 LGITWLKSSYKILISTAP 181

# RESULT 8

D97047  
 DNA ligase (NAD dependent), ligA [imported] - Clostridium acetobutylicum  
 C:Species: Clostridium acetobutylicum  
 C:Date: 14-Sep-2001 #sequence revision 14-Sep-2001 #text\_change 30-Sep-2001  
 C:Accession: D97047  
 R:Nolling, J.; Brelton, G.; Omeichenko, M.V.; Markarova, K.S.; Zeng, Q.; Gibson, R.; Lee, J.; Dally, M.J.; Bennett, G.N.; Koonin, E.V.; Smith, D.R.  
 J. Bacteriol. 183, 4823-4838, 2001  
 A:Title: Genome Sequence and Comparative Analysis of the Solvent-Producing Bacterium C.  
 A:Reference number: A96900; MUID:21359325; PMID:21359325  
 A:Accession: D97047  
 A:Status: preliminary  
 A:Molecule type: DNA  
 A:Residues: 1-663 <KUR>  
 A:Cross-references: GB:AE001437; PIDN:AAK79167.1; PID:g15024117; GSPDB:GN00168  
 A:Experimental source: Clostridium acetobutylicum ATCC824  
 C:Genetics:  
 A:Gene: CAC1195  
 C:Superfamily: polydeoxyribonucleotide synthase (NAD+)

Query Match 6.9%; Score 83; DB 2; Length 663;  
 Best Local Similarity 28.6%; Pred. No. 14;

Matches 36; Conservative 18; Mismatches 44; Indels 28; Gaps 7;

QY 102 GAVFSRVNNVLEAPFLVIGESLKGSTYNL-----LFCGSCGIPVGHLYSTHAAL 155  
 DB 368 GARVFLRRSNVLI---PEIWGVTEETBETKEIEAPTCIPYCGSEIVEGVHL----- 417  
 QY 156 AALRGHFLSSDKMVCYLLKTAIVNASEMDIQNVPLSEKIAELKEKIVLTHRLKSL 213  
 DB 418 -----FC--ENTLSCPKQWVKSIVHFASRRAMNIEGSEKTAQLEFEK--LNIKSIDL 467  
 QY 214 MKILSE 219  
 DB 468 YRTKE 473

# RESULT 9

B69099  
 formate hydrogenlyase, iron-sulfur subunit 2 - Methanobacterium thermoautotrophicum (str  
 C:Species: Methanobacterium thermoautotrophicum  
 C:Date: 05-Dec-1997 #sequence revision 05-Dec-1997 #text\_change 11-Jun-1999  
 C:Accession: B69099  
 R:Smith, D.R.; Doucette-Stamm, L.A.; Delonghery, C.; Lee, H.; Dubois, J.; Aldredge, T.;  
 Giv, D.; Spadefora, R.; Vitacore, R.; Wang, Y.; Wierzbowski, J.; Gibson, R.; Jiwani, N.  
 K.; S.; Church, G.M.; Daniele, C.J.; Mao, J.; Rice, P.; Noelling, J.; Reeve, J.N.  
 J. Bacteriol. 179, 7135-7155, 1997  
 A:Title: Complete genome sequence of Methanobacterium thermoautotrophicum Delta H: func  
 A:Reference number: A69000; MUID:98037514; PMID:9371463  
 A:Accession: B69099  
 A:Status: preliminary; nucleic acid sequence not shown; translation not shown  
 A:Molecule type: DNA  
 A:Residues: 1-143 <MTH>  
 A:Cross-references: GB:AE000929; GB:AE000666; NID:g2622853; PIDN:AA86206.1; PID:g2622853  
 A:Experimental source: strain Delta H  
 C:Genetics:  
 A:Gene: MTH1736  
 C:Superfamily: nrfC protein; ferredoxin 2[4Fe-4S] homology

Query Match 6.9%; Score 82.5; DB 2; Length 143;  
 Best Local Similarity 23.0%; Pred. No. 2.3;

Matches 35; Conservative 26; Mismatches 66; Indels 25; Gaps 6;

QY 72 PERCAVFCQCHAVLADSVHLAMDLSRLGAVF-----SRVTNNVLEAPFLVIGIEG 125  
 DB 8 PELCD--ECMKCERICPKNAIRVID-----GVPVFMHCSPPRAPCLINCPEDAIIVEVG 60  
 QY 126 SLKSGSTYNLLFCGSC--GIPVGHLYSTHAALALRGHFLSSDKMVCYLLKTAIVNAS 183  
 DB 61 AVVILIEDRCTGGLCKDACPVG--ATTNBERGVAVKCDLCTDRKFLCVWVCKGALSSES 118

QY 184 EMDIQVPLSEKIAELKEKIVLTNRKLSMK 215  
| : : : : :  
Db 119 SEDM-----MAKRDKIAGELKRLKLSMK 142

RESULT 10  
A:Accession: A85433  
C:Species: Arabidopsis thaliana (mouse-ear cress)  
C>Date: 16-Feb-2001 #sequence\_revision 16-Feb-2001 #text\_change 02-Mar-2001  
C:Accession: A85433  
R:Anonymous, The European Union Arabidopsis Genome Sequencing Consortium, The Cold Spring  
Nature 402, 769-777, 1999  
A>Title: Sequence and analysis of chromosome 4 of the plant Arabidopsis thaliana.  
A:Reference number: A85001; MUID:20083488; PMID:10617198  
A:Accession: A85433  
A>Status: preliminary  
A:Molecule type: DNA  
A:Residues: 1-493 <STO>  
A:Cross-references: GB:NC\_001268; NID:g7270615; PIDN:CAB80333.1; GSPDB:GN00140  
C:Genetics:  
A:Gene: AT4g36670  
A:Map position: 4  
C:Superfamily: glucose transport protein

Query Match 6.9%; Score 82.5; DB 2; Length 493;  
Best Local Similarity 24.9%; Pred. No. 11;  
Matches 53; Conservative 32; Mismatches 59; Indels 69; Gaps 13;

QY 79 QCAQCHAVLADSVHMLDLSRSIAVVF--SRVNNVLEAPFLVGI-----EGSLK- 128  
| : : : : :  
Db 19 QCAIVASIV--SIFGYDTGVMSGAMVFIEDLKTNDVQIEV--LTGILMLCALVGSLLA 74  
| : : : : :  
QY 129 -----GSTYNIIFGSCG--GIPVGFHL-----YSTH 152  
| : : : : :  
Db 75 GRTSDIIGRRYITVLAAILFMLGSIIMWGPNYPVLLSRTAGLGVGALMAVAPVYSAB 134  
| : : : : :  
QY 153 AALALALRG-----HFLSSDKWVCYLLK--TKAIVNAS--EMDIQVNP--LSEKIA 197  
| : : : : :  
Db 135 IATASHRGILLASPHICISITIGIITVYFSSKLEPMHIGMRLMGLAIVPSLVLAFGIL 194  
| : : : : :  
QY 198 ELKE--KIVLTNRKLSMKILSEV--TPDQSK 226  
| : : : : :  
Db 195 KMPESPRLIMQGRLEKKEKILLEVNSPEBAE 227  
| : : : : :

RESULT 11  
F87678  
DNA polymerase I [imported] - Caulobacter crescentus  
C:Species: Caulobacter crescentus  
C>Date: 20-Apr-2001 #sequence\_revision 20-Apr-2001 #text\_change 10-May-2001  
C:Accession: F87678  
R:Nietman, W.C.; Feidbylum, T.V.; Paulsen, I.T.; Nelson, K.E.; Eisen, J.; Heidelberg, J.  
B.; Laub, M.T.; DeBoy, R.T.; Dodson, R.J.; Durkin, A.S.; Gilm, M.L.; Haft, D.H.; Kolon  
B.; J.; Ermolaeva, M.; White, O.; Salzberg, S.L.; Shapir, L.; Venter, J.C.; Fraser, C.M.  
Proc. Natl. Acad. Sci. U.S.A. 98, 4136-4144, 2001  
A>Title: Complete Genome Sequence of Caulobacter crescentus.  
A:Reference number: A87249; MUID:21173698; PMID:11259647  
A:Accession: F87678  
A>Status: preliminary  
A:Molecule type: DNA  
A:Residues: 1-967 <STO>  
A:Cross-references: GB:AE005673; NID:g13425184; PIDN:AAK25426.1; GSPDB:GN00148  
C:Genetics:  
A:Gene: CC3464  
C:Superfamily: DNA-directed DNA polymerase I

Query Match 6.8%; Score 82; DB 2; Length 967;  
Best Local Similarity 24.1%; Pred. No. 29;  
Matches 51; Conservative 27; Mismatches 72; Indels 62; Gaps 11;

QY 12 CATPRGDFCGGTERAIDQ---ASFTSMEMWT---QVVKSSPLGPAGLGAEPDAPQ 65

Db 271 CDTPLPPLDALIVREDDKALAFLEQMEFRSLARVGDGSAAPGTL--DRPAPPK 328  
| : : : : :  
QY 66 LP-----SWL-----QPERC--AVPQACQCHAVLADSVHMLDLSRSIAGAVFSR 108  
| : : : : :  
Db 329 APVYVSYSYMAAARAAAHPEPVKIDHAAVACRYDLATLKAWAKATD---KGLVAFDT 384  
| : : : : :  
QY 109 VNNNVLEAPFLVIGSLKSGTYNLLFCGSCGIPVGFHLYSTHAALALRGHRCUSDK 168  
| : : : : :  
Db 385 ETD-----ALSSAT-----AGLCGV-----SLAIPGEACTYP-- 412  
| : : : : :  
QY 169 MCVYLKTKAIVNASEMDIQVPLSEKIAELK 200  
| : : : : :  
Db 413 -ISHCKADGLAFAPADPIEQIPLADVIATLK 443  
| : : : : :

RESULT 12  
D90404  
transport protein, probable [imported] - Sulfolobus solfataricus  
C:Species: Sulfolobus solfataricus  
C>Date: 24-May-2001 #sequence\_revision 24-May-2001 #text\_change 24-May-2001  
C:Accession: D90404  
R:She, Q.; Singh, R.K.; Confalonieri, F.; Zivanovic, Y.; Allard, G.; Awayez, M.J.; Chan-  
Jong, I.; Jeffries, A.C.; Kozera, C.J.; Medina, N.; Peng, X.; Thi-Ngoc, H.P.; Redder, R.  
arrett, R.A.; Ragan, M.A.; Sensen, C.W.; Van der Oost, J.  
submitted to Genbank, April 2001  
A:Description: Sulfolobus solfataricus complete genome.  
A:Reference number: A99139  
A:Accession: D90404  
A>Status: preliminary  
A:Molecule type: DNA  
A:Residues: 1-329 <KUR>  
A:Cross-references: GB:AE006641; NID:g13815639; PIDN:AAK42491.1; GSPDB:GN00155  
C:Genetics:  
A:Gene: SSO2338

Query Match 6.8%; Score 81.5; DB 2; Length 329;  
Best Local Similarity 22.2%; Pred. No. 8.2;  
Matches 40; Conservative 30; Mismatches 47; Indels 63; Gaps 7;

QY 87 IADSVHLA-WD-----LSRSIAGAVFSRVTNNV-----LE 116  
| : : : : :  
Db 33 LSESHLAVWEVFATVALPFLGRITIGSFYQVFKNSVISYCPPLGFLVITLQNLGALIF 92  
| : : : : :  
QY 117 APFLVG-----IEGSLKSGTYNLLFCGSCGIPVGFHLYSTHAALALRGHRCUS 166  
| : : : : :  
Db 93 VRFVGVIFGILLTSYAVEGAVKSGRNVALGFTTAGWPIGWV----- 134  
| : : : : :  
QY 167 DKMVCY-ILKTKAIVNASEMDIQVPLSE--KIAELKEKIVLTNRKLSMKILSEVTP 222  
| : : : : :  
Db 135 -SYAIVVLLKMNVINISGITLMLALFELNKGKRGERSKISVSFPLTISILIVSALTIP 193  
| : : : : :

RESULT 13  
B64918  
Glucuronide permease uidB - Escherichia coli (strain K-12)  
C:Species: Escherichia coli  
C>Date: 12-Sep-1997 #sequence\_revision 17-Sep-1997 #text\_change 01-Mar-2002  
C:Accession: B64918  
R:Blattner, F.R.; Plunkett III, G.; Bloch, C.A.; Perna, N.T.; Burland, V.; Riley, M.; Co  
A.; Rose, D.J.; Mau, B.; Shao, Y.  
Science 277, 1453-1462, 1997  
A>Title: The complete genome sequence of Escherichia coli K-12.  
A:Reference number: A64720; MUID:97426617; PMID:9278503  
A:Accession: B64918  
A>Status: nucleic acid sequence not shown; translation not shown  
A:Molecule type: DNA  
A:Residues: 1-457 <BLAT>  
A:Cross-references: GB:AE000257; GB:U00096; NID:g1787898; PIDN:AACT4688.1; PID:g1787902;  
A:Experimental source: strain K-12, substrain MG1655  
C:Genetics:  
A:Gene: uidB; gusB  
C:Superfamily: melibiose carrier protein

C:Keywords: carrier protein; transmembrane protein  
F:37-53/Domain: transmembrane #status predicted <TM01>  
F:81-97/Domain: transmembrane #status predicted <TM02>  
F:152-168/Domain: transmembrane #status predicted <TM03>  
F:184-200/Domain: transmembrane #status predicted <TM04>  
F:231-247/Domain: transmembrane #status predicted <TM05>  
F:263-279/Domain: transmembrane #status predicted <TM06>  
F:310-326/Domain: transmembrane #status predicted <TM07>  
F:408-424/Domain: transmembrane #status predicted <TM08>

Query Match 6.8%; Score 81.5; DB 2; Length 457;  
Best Local Similarity 25.2%; Pred. No. 12;

Matches 37; Conservative 30; Mismatches 53; Indels 27; Gaps 8;

QY 95 WDLRSIGAVVFSRVTVNNV-----VLEA-----PFIWGIEGSLKSTYNLL-FCGSCG 141  
DB 317 WSLPVALVALALASIGQVITWMMALPADVEYGEYLTGR--IEGLTSLFSTRKCG 374  
QY 142 IPVGPHLYSTHAALALRGHFC--LSSDKWVCYLKTKAIVNASMDIQNV-----PLSE 194  
DB 375 QAIG---GSIPAFILGSGYIANQVQTPREVIMGIRTSIALVPCGFMILAFVITWFFPLTD 431  
QY 195 KIAELKEKIVLTNRKLSMKILSEVT 221  
DB 432 K--KFEKIVEIDNRKKVQQQLISDIT 456

## RESULT 14

G90919 glucuronide permease [imported] - Escherichia coli (strain O157:H7, substrain RIMD 05099

C:Species: Escherichia coli  
C>Date: 18-Jul-2001 #sequence\_revision 18-Jul-2001 #text\_change 03-Aug-2001

C:Accession: C90919

R:Hayashi, T.; Makino, K.; Onishi, M.; Kurokawa, K.; Ishii, K.; Yokoyama, K.; Han, C.G. gasawara, N.; Yasunaga, T.; Kuhara, S.; Shiba, T.; Hattori, M.; Shingawa, H.

DNA Res. 8, 11-22, 2001

A:Title: Complete genome sequence of enterohemorrhagic Escherichia coli O157:H7 and geno

A:Reference number: A99629; MUID:21156231; PMID:11258796

A:Accession: C90919

A:Status: preliminary

A:Molecule type: DNA

A:Residues: 1-457 <HAY>  
A:Cross-references: GB:BA000007; PIDN:BA03746.1; PID:g13361790; GSPDB:GN00154

A:Experimental source: strain O157:H7, substrain RIMD 0509952

C:Genetics:

A:Gene: BGS2323

C:Superfamily: melibiose carrier protein

Query Match 6.8%; Score 81.5; DB 2; Length 457;  
Best Local Similarity 25.2%; Pred. No. 12;

Matches 37; Conservative 30; Mismatches 53; Indels 27; Gaps 8;

QY 95 WDLRSIGAVVFSRVTVNNV-----VLEA-----PFIWGIEGSLKSTYNLL-FCGSCG 141  
DB 317 WSLPVALVALALASIGQVITWMMALPADVEYGEYLTGR--IEGLTSLFSTRKCG 374  
QY 142 IPVGPHLYSTHAALALRGHFC--LSSDKWVCYLKTKAIVNASMDIQNV-----PLSE 194  
DB 375 QAIG---GSIPAFILGSGYIANQVQTPREVIMGIRTSIALVPCGFMILAFVITWFFPLTD 431  
QY 195 KIAELKEKIVLTNRKLSMKILSEVT 221  
DB 432 K--KFEKIVEIDNRKKVQQQLISDIT 456

## RESULT 15

H85767

glucuronide permease [imported] - Escherichia coli (strain O157:H7, substrain EDL933)

C:Species: Escherichia coli

C>Date: 16-Feb-2001 #sequence\_revision 16-Feb-2001 #text\_change 14-Sep-2001

C:Accession: H85767

R:Perera, N.T.; Plummett III, G.; Burland, V.; Mau, B.; Glasner, J.D.; Rose, D.J.; Mayhew

Iller, L.; Grotbeck, E.J.; Davis, N.W.; Lim, A.; Dimalanta, E.; Potamoustis, K.; Apodaca,

Nature 409, 529-533, 2001

A:Title: Genome sequence of enterohemorrhagic Escherichia coli O157:H7.

A:Reference number: A85480; MUID:21074935; PMID:11206551

A:Accession: H85767

A:Status: preliminary

A:Molecule type: DNA

A:Residues: 1-457 <SNO>

A:Cross-references: GB:AE005174; NID:g12515601; PIDN:AAG56604.1; GSPDB:GN00145; UWGP:Z2

A:Experimental source: strain O157:H7, substrain EDL933

C:Genetics:

A:Gene: uidB

C:Superfamily: melibiose carrier protein

Query Match 6.8%; Score 81.5; DB 2; Length 457;  
Best Local Similarity 25.2%; Pred. No. 12;

Matches 37; Conservative 30; Mismatches 53; Indels 27; Gaps 8;

QY 95 WDLRSIGAVVFSRVTVNNV-----VLEA-----PFIWGIEGSLKSTYNLL-FCGSCG 141  
DB 317 WSLPVALVALALASIGQVITWMMALPADVEYGEYLTGR--IEGLTSLFSTRKCG 374  
QY 142 IPVGPHLYSTHAALALRGHFC--LSSDKWVCYLKTKAIVNASMDIQNV-----PLSE 194  
DB 375 QAIG---GSIPAFILGSGYIANQVQTPREVIMGIRTSIALVPCGFMILAFVITWFFPLTD 431  
QY 195 KIAELKEKIVLTNRKLSMKILSEVT 221  
DB 432 K--KFEKIVEIDNRKKVQQQLISDIT 456

Search completed: July 20, 2004, 11:18:11

Job time : 18 secs



GenCore version 5.1.6  
Copyright (c) 1993 - 2004 Compugen Ltd.

OM protein - protein search, using sw model

Run on: July 20, 2004, 11:13:59 ; Search time 40 Seconds  
(without alignments)  
1806.342 Million cell updates/sec

Title: US-09-942-052a-728  
Perfect score: 1198  
Sequence: 1 MAQPLRRSRRCATPRGRDF.....LKSIMKILSEVTPDQSKPEN 229

Scoring table: BLOSUM62  
Gapop 10.0 , Gapext 0.5

Searched: 1017041 seqs, 315518202 residues

Total number of hits satisfying chosen parameters: 1017041

Minimum DB seq length: 0  
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%  
Maximum Match 100%  
Listing first 45 summaries

Database :

SPTREMBL\_25.\*  
1: sp archaea.\*  
2: sp bacteria.\*  
3: sp fungi.\*  
4: sp human.\*  
5: sp invertebrate.\*  
6: sp mammal.\*  
7: sp mhc.\*  
8: sp organelle.\*  
9: sp phage.\*  
10: sp plant.\*  
11: sp rodent.\*  
12: sp virus.\*  
13: sp vertebrate.\*  
14: sp unclassified.\*  
15: sp viirus.\*  
16: sp bacteriap.\*  
17: sp archaeap.\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

# SUMMARIES

Result No.	Score	Query Match	Length	DB	ID	Description
1	122	10.2	218	11	Q9CXR6	Q9cxi6 mus musculu
2	115.5	9.6	204	11	Q9CZU6	Q9czi6 mus musculu
3	92	7.7	361	10	Q8S611	Q8s611 oryza sativ
4	92	7.7	695	16	Q9CH87	Q9ch87 lactococcus
5	91	7.6	652	4	Q96EP1	Q96ep1 homo sapien
6	90.5	7.6	597	16	Q83693	Q83693 treponema p
7	89	7.4	306	4	Q9N32	Q9n32 homo sapien
8	89	7.4	623	4	Q9NVD5	Q9nvd5 homo sapien
9	89	7.4	652	4	Q96SL3	Q96sl3 homo sapien
10	89	7.3	664	4	Q9NRT4	Q9nrt4 homo sapien
11	87.5	7.3	663	11	Q8BJZ9	Q8bjz9 mus musculu
12	87	7.3	664	11	Q810L3	Q810l3 mus musculu
13	86.5	7.2	669	16	Q97F05	Q97f05 clostridium
14	86	7.2	145	16	Q839T4	Q839t4 enterococcu
15	85	7.1	337	16	Q82HU1	Q82hu1 streptomyce
16	85	7.1	671	5	Q815K5	Q815k5 plasmodium

17	85	7.1	2194	10	Q40360	Q40360 medicago sa
18	84.5	7.1	361	16	Q8YV34	Q8yv34 anabena sp
19	84.5	7.1	740	4	Q81WD2	Q81wd2 homo sapien
20	84	7.0	496	4	Q96SL8	Q96sl8 homo sapien
21	84	7.0	1129	16	Q7UKD4	Q7ukd4 rhodopirell
22	84	7.0	1947	10	Q84SE8	Q84se8 oryza sativ
23	83.5	7.0	800	9	Q7J5U5	Q7j5u5 bacterioph
24	83	6.9	475	10	Q7XQK5	Q7xqk5 oryza sativ
25	83	6.9	579	12	Q8BEN6	Q8ben6 callitrich
26	83	6.9	663	16	Q97J58	Q97j58 clostridium
27	82.5	6.9	143	17	Q27769	Q27769 methanobact
28	82.5	6.9	423	16	Q87A11	Q87a11 xylella fas
29	82.5	6.9	493	10	Q23213	Q23213 arabidopsis
30	82.5	6.9	640	12	Q56294	Q56294 human herpe
31	82	6.8	697	16	Q9A2U2	Q9a2u2 caulobacter
32	82	6.8	2216	10	Q9LV03	Q9lv03 arabidopsis
33	81.5	6.8	329	17	Q97W97	Q97w97 sulfolobus
34	81.5	6.8	457	16	Q8X673	Q8x673 escherichia
35	81.5	6.8	563	11	Q99KX1	Q99kx1 mus musculu
36	81.5	6.8	672	11	Q80V27	Q80v27 mus musculu
37	81.5	6.8	2172	2	Q7X4R4	Q7x4r4 streptomyce
38	81	6.8	418	13	Q7ZYX0	Q7zyx0 brachydanio
39	81	6.8	443	10	Q8LPK8	Q8lpk8 arabidopsis
40	81	6.8	1697	12	Q8QZP9	Q8qzp9 crimean-con
41	80.5	6.7	766	10	Q23161	Q23161 arabidopsis
42	80.5	6.7	768	10	Q8VZC5	Q8vzc5 arabidopsis
43	80	6.7	105	16	Q8FNE5	Q8fne5 corynebacte
44	80	6.7	431	16	Q88U19	Q88u19 lactobacill
45	80	6.7				

## ALIGNMENTS

RESULT 1  
Q9CXR6 PRELIMINARY: PRT: 218 AA.  
AC Q9CXR6; 01-JUN-2001 (TREMBLrel. 17, Created)  
DT 01-JUN-2001 (TREMBLrel. 17, Last sequence update)  
DT 01-JUN-2001 (TREMBLrel. 17, Last annotation update)  
DE 2610039C10R1k protein.  
GN 2610039C10R1K.  
OS Mus musculus (Mouse).  
OC Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi;  
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Mus.  
OX NCBI\_TaxId=10090;  
RN [1]  
RP SEQUENCE FROM N.A.  
RC STRAIN=C57BL/6J; TISSUE=Embryo;  
RX MEDLINE=21085660; PubMed=11217851;  
RA Kawai J., Shingawa A., Shibata K., Yoshino M., Itoh M., Ishii Y.,  
RA Aizawa K., Izawa M., Nishi K., Kiyosawa H., Kondo S., Yamataka I.,  
RA Saito T., Okazaki Y., Gojobori T., Bono H., Kasukawa T., Saito R.,  
RA Kadota K., Matsuda H.A., Ashburner M., Batilov S., Casavant T.,  
RA Fleschmann W., Gaasterland T., Gissi C., King B., Kochia R.,  
RA Kuehl P., Lewis S., Matsuo Y., Nikaide I., Pesole G., Quackenbush J.,  
RA Schriml L.M., Staubli F., Suzuki R., Tomita M., Wagner L., Washio T.,  
RA Sakai K., Okide T., Furuno M., Aono H., Baldarelli R., Barsh G.,  
RA Blake J., Boffelli D., Bojunga N., Carninci P., de Bonaldo M.F.,  
RA Brownstein M.J., Bult C., Fletcher C., Fujita M., Gariboldi M.,  
RA Gustincich S., Hill D., Hofmann M., Hume D.A., Kamuya M., Lee N.H.,  
RA Lyons P., Marchionni L., Mashima J., Mazzarelli J., Mombarts P.,  
RA Mordone P., Ring B., Ringwald M., Rodriguez I., Sakamoto N.,  
RA Sasaki H., Sato K., Schoenbach C., Seya T., Shibata Y., Storch K.-F.,  
RA Suzuki H., Toyo-Oka K., Wang K.H., Weitz C., Whitaker C., Wilming L.,  
RA Wyszynaw-Boris A., Yoshida K., Hasegawa Y., Kawai H., Kohseki S.,  
RA Hayashizaki Y.;  
RT "Functional annotation of a full-length mouse cDNA collection.";  
RL Nature 409:685-690(2001).  
DR EMBL, AK014084; BAB29147.1; -;  
DR MGD; WGI:1913828; 2610039C10R1k.

SQ SEQUENCE 218 AA: 24447 MW: 553305BHPDDE8C2 CRC64;  
 Query Match 10.2%; Score 122; DB 11; Length 218;  
 Best Local Similarity 23.3%; Pred. No. 0.00087;  
 Matches 48; Conservative 30; Mismatches 102; Indels 26; Gaps 6;  
 QY 29 DASFTTSMWDTQYVKGSSPLGPAGLGAEEPAAGPQLPSWLOPERCAVFOCAQCHAVLA 88  
 DB 15 DSRRIYRLQKMAN--MSADALGLEKEERPEEFAAAE--LDFLCARCRPLG 64  
 QY 89 DSVHLAMDLSR-SLGAIVVRSYRTNNVLAEPFLVIGSLKSTYNTLFFCGSCGIPVGFH 147  
 DB 65 DS--LTVVASQGDPTNCILRSVSCNVSDKEPKLSKCDDEDCILEALYTCGCSISLGY 122  
 QY 148 LYSTAAALALRGHRCCLSSDKKVCYLL-----KTKAIVNASEMDIQVPLSEKIAE 198  
 DB 123 YRCTPKNDPKYKDLCSLVEVEESTYLLGSSEKQIVSKDKEFNLE---SVELEKSKIQ 178  
 QY 139 LKEKIVLTHNRKLSMKLITSEVTPQ 224  
 DB 179 MEVVTALQKRLREVSKEKLSIAQPEQ 204  
 RESULT 2  
 Q9CZJ6 PRELIMINARY; PRT: 204 AA.  
 AC Q9CZJ6;  
 DT 01-0UN-2001 (TRENBLrel. 17, Created)  
 DT 01-0UN-2001 (TRENBLrel. 17, Last sequence update)  
 DE 01-0UN-2001 (TRENBLrel. 17, Last annotation update)  
 GN 2610039C10RIK protein.  
 OS 2610039C10RIK  
 OS Mus musculus (Mouse).  
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
 OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.  
 OX NCBI\_TaxID=10090;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RC STRAIN=C57BL/6J; TISSUE=Embryo;  
 RX MEDLINE=21085660; PubMed=11217851;  
 RA Kawai J., Shiragawa A., Shibata K., Yoshino M., Itoh M., Ishii Y.,  
 RA Arakawa T., Hara A., Fukunishi Y., Kono H., Adachi J., Fukuda S.,  
 RA Aizawa K., Izawa M., Nishi K., Kiyosawa H., Kondo S., Yamataka I.,  
 RA Saito T., Okazaki Y., Gojohori T., Bono H., Kasukawa T., Saito R.,  
 RA Kadota K., Matsuda H.A., Ashburner M., Batalov S., Casavant T.,  
 RA Fleischman W., Gaasterland T., Gissi C., King B., Kochiwa H.,  
 RA Kuehl P., Lewis S., Matsuo Y., Nikaido I., Pesole G., Quackenbush J.,  
 RA Schiomi L.M., Stenbli F., Suzuki R., Tomita M., Wagner L., Washio T.,  
 RA Sakai K., Okido T., Furuno M., Aono H., Baldarelli R., Barsh G.,  
 RA Blake J., Boffelli D., Bojunga N., Carninci P., de Bonaldo M.F.,  
 RA Brownstein M.-J., But C., Fletcher C., Fujita M., Gariboldi M.,  
 RA Gustincich S., Hill D., Hofmann M., Hume D.A., Kamiya M., Lee N.H.,  
 RA Lyons P., Matchonni L., Mashima J., Mazzarelli J., Monbaerts P.,  
 RA Nordone P., Ring B., Ringwald M., Rodriguez I., Sakamoto N.,  
 RA Sasaki H., Sato K., Schoenbach C., Seya T., Shibata Y., Storch K.-F.,  
 RA Suzuki H., Toyooka K., Wang K.H., Weitz C., Whitlaker C., Wilming L.,  
 RA Wyrshaw-Boris A., Yoshida K., Hasegawa Y., Kawaji H., Kohetsuki S.,  
 RA Hayaishizaki Y.;  
 RT "Functional annotation of a full-length mouse cDNA collection."  
 RL Nature 409:685-690(2001).  
 DR EMBL; AK012533; BAB28302.1; -  
 DR MGI; MGI:1913828; 2610039C10RIK.  
 SQ SEQUENCE 204 AA; 22949 MW; BBD667250FDF9B0F CRC64;  
 Query Match 9.6%; Score 115.5; DB 11; Length 204;  
 Best Local Similarity 25.6%; Pred. No. 0.0035;  
 Matches 50; Conservative 29; Mismatches 97; Indels 19; Gaps 7;  
 QY 29 DASFTTSMWDTQYVKGSSPLGPAGLGAEEPAAGPQLPSWLOPERCAVFOCAQCHAVLA 88  
 DB 15 DSRRIYRLQKMAN--MSADALGLEKEERPEEFAAAE--LDFLCARCRPLG 64  
 QY 89 DSVHLAMDLSR-SLGAIVVRSYRTNNVLAEPFLVIGSLKSTYNTLFFCGSCGIPVGFH 147  
 DB 65 DS--LTVVASQGDPTNCILRSVSCNVSDKEPKLSKCDDEDCILEALYTCGCSISLGY 122  
 QY 148 LYSTAAALALRGHRCCLSSDKKVCYLL-----KTKAIVNASEMDIQVPLSEKIAE 198  
 DB 123 YRCTPKNDPKYKDLCSLVEVEESTYLLGSSEKQIVSKDKEFNLE---SVELEKSKIQ 178  
 QY 139 LKEKIVLTHNRKLSMKLITSEVTPQ 224  
 DB 179 MEVVTALQKRLREVSKEKLSIAQPEQ 204

```

Db      65 DS--LTTWASQGDNTNCTILRSVCSNVSVKPEPLSKCRDDEGCTILALCTGSGSLGV 122
QY      148 LYSTHAALALFGHRLSSDRKVCYTL--KTAIVASEMDIQNVPLSEKIALKEKIYI 205
Db      123 YRCTENKLDYKDFCLCSVEAVESYTLGSGSEKQIV--SEDKELFNL---ESRVEIKSIK 178
QY      206 TNNRLKSLMKILSEV 220
Db      179 MEVLTAALQCKLREV 193

RESULT 3
08S611
ID      08S611      PRELIMINARY;      PRT;      361 AA.
AC      08S611:
DT      01-JUN-2002 (TREMBlrel. 21, Created)
DT      01-JUN-2002 (TREMBlrel. 21, Last sequence update)
DT      01-OCT-2003 (TREMBlrel. 25, Last annotation update)
DE      Putative gypsy-type retrotransposon protein.
GN      OSUNBA009652.5
OS      Oryza sativa (japonica cultivar-group).
OC      Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
OC      Spermatophyta; Magnoliopsida; Liliopsida; Poales; Poaceae;
OC      Ehrhartoideae; Oryzaceae; Oryza.
OX      NCBI_TaxID=39947;
RN      RN
RP      [1]
RP      SEQUENCE FROM N.A.
RA      Wang R.A., Yu Y., Soderlund C., Chen M., Kim H.-R., Rambo T.,
RA      Saeki C., Henry D., Oates R., Simmons J.;
RT      "Rice Genomic Sequence.";
RL      Submitted (MAR-2002) to the EMBL/GenBank/DBJ databases.
RN      RN
RP      [2]
RP      SEQUENCE FROM N.A.
RA      STRAIN=cv. Nipponbare;
RA      The Rice Chromosome 10 Sequencing Consortium;
RT      "In-depth view of structure, activity, and evolution of rice
RT      chromosome 10.";
RL      Science 300:1566-1569(2003).
RN      RN
RP      [3]
RP      SEQUENCE FROM N.A.
RA      STRAIN=cv. Nipponbare;
RA      Buehl C.R., Wing R.A., McCombie W.R., Messing J., Yuan Q.;
RL      Submitted (MAY-2003) to the EMBL/GenBank/DBJ databases.
DR      EMBL; AC099400; AAL91599.1; -
DR      EMBL; AB017047; AAP51763.1; -.
DR      Genemex; Q8S611; -.
SQ      SEQUENCE 361 AA; 38805 MW; F883BB3E8B8FEFF45 CRC64;

Query Match      7.7%; Score 92; DB 10; Length 361;
Best Local Similarity 25.3%; Pred. No. 1.6;
Matches 61; Conservative 28; Mismatches 92; Indels 60; Gaps 14
QY      2 AAQPLRRSRSCAPPKRDPCGGTERAIDQASFTTSMEMDTQVYVKGSSPLGAGAEERPA 61
Db      49 APTPLPRRAVRAKAAODGSGGTTSSAPAVAST-----DVVVVGSRSEATPSG--PASDPV 102
QY      62 AGPQLP-----SW--IQPERCAVPGCAQCHAVLADSVLAMDLSLSLAQVVSRYTNNVYL 115
Db      103 AGKSPAAVATSWBELQVWGRRLLEAGA--RVIGREIAEAGLEHRM-----SELGN--- 152
QY      116 EAPFLVIGESLKGSTNYL--FCGSGCI-----PVGFHYSTHAALAT-----RGH 161
Db      153 ----LSRIKRSLSR-VYITGLHQLAGKKGISITIPANDDEBSLTSLIAELAAAMEITPSKH 207
QY      162 FCLSSDRM-----VCYLTKTAIVYVASEMDIQNV-----PLSEKIAELYKBI 203
Db      208 AARIGEMSNRIYTGACHILACVRIAH-PEIDLRITLIDQEBASDARKDQVMEVEVDLCKSV 266
QY      204 V 204
Db      267 L 267

```

## RESULT 4

09CH87 PRELIMINARY; PRT; 695 AA.

ID 09CH87

AC 09CH87

DT 01-JUN-2001 (TREMblrel. 17, Created)

DT 01-JUN-2001 (TREMblrel. 17, Last sequence update)

DT 01-OCT-2003 (TREMblrel. 25, Last annotation update)

DE Copper-potassium transporting ATPase B.

GN COB OR L10851.

OS Lactococcus lactis (subsp. lactis) (Streptococcus lactis).

OC Bacteria; Firmicutes; Lactobacillales; Streptococcaceae; Lactococcus.

NCBI\_Taxid=1360;

SEQUENCE FROM N.A.

RC STRAIN=114403;

RA Bojotin A., Winkler P., Manger S., Jallion O., Malarme K., Weissenbach J., Ehrlich S.D., Sorokin A.;

RT The complete genome sequence of the lactic acid bacterium Lactococcus lactis ssp. lactis 114403 ";

RL Genome Res. 11:731-753(2001).

EMBL; AE006319; AAK04949.1; --

DR PIR; C86731; C86731.

DR GO; GO:0016021; C:integral to membrane; IEA.

DR GO; GO:0005524; F:ATPase binding; IEA.

DR GO; GO:0015662; F:ATPase activity, coupled to transmembrane m. . .; IEA.

DR GO; GO:0016787; F:hydrolase activity; IEA.

DR GO; GO:0008152; P:metabolism; IEA.

DR GO; GO:0030001; P:metal ion transport; IEA.

DR InterPro; IPR006403; ATPase-IB1 Cu.

DR InterPro; IPR006416; ATPase-IB1 hv.

DR InterPro; IPR001257; ATPase-IB1-E2.

DR InterPro; IPR008250; E1-E2 ATPase-reg.

DR InterPro; IPR005834; Hydrolase.

DR Pfam; PF00122; E1-E2 ATPase; 1.

DR Pfam; PF00702; Hydrolase; 1.

DR PRINTS; PR00119; CATATPASE.

DR TIGRfams; TIGR01511; ATPase-IB1 Cu; 1.

DR TIGRfams; TIGR01525; ATPase-IB1 hv; 1.

DR TIGRfams; TIGR01494; ATPase-P-type; 2.

DR PROSITE; PS00154; ATPASE\_E1\_E2; 1.

KW Complete proteome.

SEQUENCE 695 AA; 75981 MW; 1B4947C32A0FA0F CRC64;

Query Match 7.7%; Score 92; DB 16; Length 695;

Best Local Similarity 25.3%; Pred. No. 3.7;

Matches 38; Conservative 28; Mismatches 50; Indels 34; Gaps 7;

DB 101 LGAVVSRVTNNVLEAPFLVIGESLKGSTVYLFCGSCGIPVGFHLYSTHALLALRG 160

101 LGTILF-----FYSGTPPPSGAKGLSKRKRPMMLITMGITVAI-AISVAATIMSNG 153

QY 161 HF-----CLSDKMYCYLLTKTAIVNASE-MDI-----QNVPLSE-K 195

DB 154 HWGMNFWFELATLIVMLTGLHLEMAKIMGAGDALDLASLVKKAHLKSGKVEHSEIK 213

QY 196 IAELEKRYITNHRKLSIMKILSEVPDGS 225

DB 214 VGDLL-LVYENKRIADGLILSEALVDS 241

RESULT 5

096EP1 PRELIMINARY; PRT; 652 AA.

ID 096EP1

AC 096EP1

DT 01-DEC-2001 (TREMblrel. 19, Created)

DT 01-DEC-2001 (TREMblrel. 19, Last sequence update)

DT 01-OCT-2003 (TREMblrel. 25, Last annotation update)

DE Hypothetical protein.

OS Homo sapiens (human).

OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Primates; Catarrhini; Hominiidae; Homo.

NCBI\_Taxid=9606;

SEQUENCE FROM N.A.

RA Tissue=Placenta;

RA Strausberg R.;

RL Submitted (AUG-2001) to the EMBL/GenBank/DBJ databases.

CC -; SIMILARITY: CONTAINS 1 RING-TYPE ZINC FINGER.

DR EMBL; BC012072; AAH12072.1; --

DR InterPro; IPR000253; FHA.

DR InterPro; IPR008984; SMAD\_FHA.

DR InterPro; IPR001841; Znf\_Ring.

DR Pfam; PF00498; FHA; 1.

DR Pfam; PF00097; ZF-C3HC4; 1.

DR SMART; SM00240; FHA; 1.

DR SMART; SM00184; RING; 1.

DR PROSITE; PS50006; FHA DOMAIN; 1.

DR PROSITE; PS00518; ZF\_RING\_1; 1.

DR PROSITE; PS50089; ZF\_RING\_2; 1.

DR Hypothetical protein; Metal-binding; Zinc; Zinc-finger.

KW Sequence 652 AA; 72058 MW; 1A50773CEA679F07 CRC64;

Query Match 7.6%; Score 91; DB 4; Length 652;

Best Local Similarity 23.5%; Pred. No. 4.3;

Matches 53; Conservative 27; Mismatches 72; Indels 74; Gaps 12;

QY 8 HRSRCATPRDPC-----GTERALDQ-----SFTSMEMTQVYKGS-----S 48

DB 422 YRRQAQPPH---CPAPRGEPAPQALDADSTSVSLTTAVADYVCPQSGHALTCGFRQ 478

QY 49 PLGPRGLGAE-EPAAGPOLPMLQPERCAVQCACCHVILVDSVHLANDLSR----- 99

DB 479 PMPDRVREREDPRVAPQ-----QCAVC---LDPFCHLYMGCTRTGCYCGLA 522

QY 100 -----SLGAVVSRVTNNVLEAPFLVIGESLKGSTY-NLIFCGSCGIPVGFHLYSTHA 153

DB 523 PCPEINLGDKCLDGLVANNSTYESDLTKNYIAT-RLGLTKMNL-----TFS 566

QY 154 ALAALRGHFCIS-----SDKMYCYLLTKTAIVNASEMDIQNVPLSE 194

DB 567 LVALQGVFLSDYRVTGDTVLYCCGLRSPRELTYYQYRQNPASH 612

RESULT 6

083693 PRELIMINARY; PRT; 597 AA.

ID 083693

AC 083693

DT 01-NOV-1998 (TREMblrel. 08, Created)

DT 01-NOV-1998 (TREMblrel. 08, Last sequence update)

DT 01-JUN-2003 (TREMblrel. 24, Last annotation update)

DE Phosphoribosylglycinamide formyltransferase, putative.

GN TP0695.

OS Treponema pallidum.

OC Bacteria; Spirochaetes; Spirochaetales; Spirochaetaceae; Treponema.

NCBI\_Taxid=160;

SEQUENCE FROM N.A.

RC STRAIN=Nichols;

RA Fraser C.M., Norris S.J., Weinstock G.M., White O., Sutton G.G., Dodson R., Gwinn M., Hickey E.K., Clayton R., Ketchum K.A., Sodergren E., Hardam J.M., McLeod M.P., Salzberg S., Peterson J., Khalak H., Richardson D., Howell J.K., Chidambaram M., Utterback T., McDonald L., Artlich P., Bowman C., Colton M.D., Fujii C., Gattland S., Hatch B., Horst K., Roberts K., Sandusky M., Weidman J., Smith H.O., Venter J.C.;

RA "Complete genome sequence of Treponema pallidum, the syphilis spirochete.";

RT Science 281:375-388(1998).

RL EMBL; AE001243; AAC65662.1; --

DR PIR; D71293; D71293.

DR TIGR; TP0695; --

DR GO; GO:0016740; F:transferase activity; IEA.

DR InterPro; IPR003135; ATP-grasp.

DR Pfam; PF02222; ATP-grasp; 1.  
 KW Transferase; Complete proteome.  
 SQ SEQUENCE 597 AA; 63325 MW; 7875117C9BFD6a6 CRC64;

Query Match  
 Best Local Similarity 7.6%; Score 90.5; DB 16; Length 597;  
 Matches 51; Conservative 19; Mismatches 47; Indels 65; Gaps 12;

QY 12 CATPRGDFCGGTRRAID-----QASFT-----TSMEMPTQVYKSSPLGPA 53  
 Db 90 CALP--GHRLEATKATADKTRMRACFTRARLCRFRFFLEPDSFAMDT-----PPGHA 140  
 QY 54 GLGAEPPAAGPQLPSWLP-ERCAVFOC--AOCHAVLADSVHLAMDLSRSIGAVVFSRV 110  
 Db 141 RLCSHLHSAGISFPLVVKPTDNMGARGCTIAQCKDTLNNCAVARQPSRS----- 190  
 QY 111 NNVLAPLPLVIGESLKGSTYVNLFCGSCGIPVGFHLYSTHAAALRGHFCLSDRMV 170  
 Db 191 GRVYII-FTLVGREPSLEG---LIFDGT-----LYVT--ALA-----DRHI 225  
 QY 171 CY 172  
 Db 226 CF 227

## RESULT 7

Q9NT32 PRELIMINARY; PRT; 306 AA.

AC Q9NT32;  
 DT 01-OCT-2000 (TREMBLrel. 15, Created)  
 DT 01-OCT-2000 (TREMBLrel. 15, Last sequence update)  
 DT 01-JUN-2003 (TREMBLrel. 24, Last annotation update)  
 DE Hypothetical protein (Fragment).  
 GN DKFZP434N2420.  
 OS Homo sapiens (Human).  
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
 OC Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.  
 OX NCBI\_TaxID=9606;  
 RN (1)  
 RP SEQUENCE FROM N.A.  
 RC TISSUE=Testis;  
 RA Othenwelder B., Obermaier B., Mewes H.W., Gassenhuber J., Wiemann S.,  
 RL Submitted (JAN-2000) to the EMBL/GenBank/DBJ databases.  
 DR EMBL; AL137561; CAB70812.1; -.  
 DR PIR; T46399; T46399.  
 KW Hypothetical protein.  
 FT NON TER  
 SQ SEQUENCE 306 AA; 34501 MW; 6B50F04601FB2939 CRC64;

Query Match  
 Best Local Similarity 7.4%; Score 89; DB 4; Length 306;  
 Matches 52; Conservative 26; Mismatches 74; Indels 74; Gaps 11;

QY 8 HSRRCATPRRGDFC-----GTERAIDQA-----SFTSMEMPTQVYKSSPL----- 50  
 Db 76 YRQAQPH---CPABEGEPAPQALGAPSTSVLTAVDYPVPLQSHALCTCCFQ 132  
 QY 51 ---GPAGLGAEPAPGQPLPSWLQPERCAVFOCAOCHAVLADSVHLAMDLSR----- 99  
 Db 133 PMPDRRAEEDQPRVAPQ-----QCAVC---LQPFCHLYWCTRTGCGCIA 176  
 QY 100 -----SLGAVFSRVYNNVLEAPPLVIGESLKGSTY-NLIFCGSCGIPVGFHLYSTHA 153  
 Db 177 PFCEINLGDCKIDGVILNNNSYESDILKNYLAT-RGLTWKNNL-----TES 220  
 QY 154 ALAALRGHFCIS-----SDKAVCYILKTKAIVNASEMDIQNVPLSE 194  
 Db 221 LVALLQGVFLISDVRVGTDTVLCYCCGRLRSFRELTYQYQONIPASE 266

RESULT 8  
 Q9NVDS PRELIMINARY; PRT; 623 AA.

AC Q9NVDS;  
 DT 01-OCT-2000 (TREMBLrel. 15, Created)  
 DT 01-OCT-2000 (TREMBLrel. 15, Last sequence update)  
 DT 01-OCT-2003 (TREMBLrel. 25, Last annotation update)  
 DE Hypothetical protein FLJ10796.  
 OS Homo sapiens (Human).  
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
 OC Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.  
 OX NCBI\_TaxID=9606;  
 RN (1)  
 RP SEQUENCE FROM N.A.  
 RA Isegai T., Ota T., Hayashi K., Sugiyama T., Otsuki T., Suzuki Y.,  
 RA Nishikawa T., Nagai K., Sugano S., Aotsuka S., Yoshikawa Y.,  
 RA Matsunawa H., Ishii S., Kawai Y., Saito K., Yamamoto J., Wakamatsu A.,  
 RA Nakamura Y., Nagahari K., Masuho Y., Sasaki N.,  
 RT "NEDO human cDNA sequencing project."  
 RL Submitted (FEB-2000) to the EMBL/GenBank/DBJ databases.  
 CC -1- SIMILARITY: CONTAINS 1 RING-TYPE ZINC FINGER.  
 DR EMBL; AK001658; BA91817.1; -.  
 DR InterPro; IPR000253; FHA.  
 DR InterPro; IPR008984; SMAD\_FHA.  
 DR InterPro; IPR001841; Znf\_Ring.  
 DR Pfam; PF00498; FHA; 1.  
 DR Pfam; PF00097; zf-C3HC4; 1.  
 DR SMART; SM00240; FHA; 1.  
 DR SMART; SM00184; RING; 1.  
 DR PROSITE; PS50006; FHA DOMAIN; 1.  
 DR PROSITE; PS00518; ZF\_RING\_1; 1.  
 DR PROSITE; PS50089; ZF\_RING\_2; 1.  
 KW Hypothetical protein\_Metal-binding; Zinc; Zinc-finger.  
 SQ SEQUENCE 623 AA; 69204 MW; 45186D3DA52711 CRC64;

Query Match  
 Best Local Similarity 7.4%; Score 89; DB 4; Length 623;  
 Matches 52; Conservative 26; Mismatches 74; Indels 74; Gaps 11;

QY 8 HSRRCATPRRGDFC-----GTERAIDQA-----SFTSMEMPTQVYKSSPL----- 50  
 Db 393 YRQAQPH---CPABEGEPAPQALGAPSTSVLTAVDYPVPLQSHALCTCCFQ 449  
 QY 51 ---GPAGLGAEPAPGQPLPSWLQPERCAVFOCAOCHAVLADSVHLAMDLSR----- 99  
 Db 450 PMPDRRAEEDQPRVAPQ-----QCAVC---LQPFCHLYWCTRTGCGCIA 493  
 QY 100 -----SLGAVFSRVYNNVLEAPPLVIGESLKGSTY-NLIFCGSCGIPVGFHLYSTHA 153  
 Db 494 PFCEINLGDCKIDGVILNNNSYESDILKNYLAT-RGLTWKNNL-----TES 537  
 QY 154 ALAALRGHFCIS-----SDKAVCYILKTKAIVNASEMDIQNVPLSE 194  
 Db 538 LVALLQGVFLISDVRVGTDTVLCYCCGRLRSFRELTYQYQONIPASE 583

## RESULT 9

Q96SL3 PRELIMINARY; PRT; 652 AA.

AC Q96SL3;  
 DT 01-DEC-2001 (TREMBLrel. 19, Created)  
 DT 01-DEC-2001 (TREMBLrel. 19, Last sequence update)  
 DT 01-OCT-2003 (TREMBLrel. 25, Last annotation update)  
 DE Hypothetical protein FLJ14781.  
 OS Homo sapiens (Human).  
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
 OC Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.  
 OX NCBI\_TaxID=9606;  
 RN (1)  
 RP SEQUENCE FROM N.A.  
 RA Isegai T., Ota T., Hayashi K., Sugiyama T., Otsuki T., Suzuki Y.,  
 RA Nishikawa T., Nagai K., Sugano S., Aotsuka S., Yoshikawa Y.,  
 RA Matsunawa H., Ishii S., Kawai Y., Saito K., Yamamoto J., Wakamatsu A.,  
 RA Nakamura Y., Nagahari K., Masuho Y., Sasaki N.,  
 RT "NEDO human cDNA sequencing project."  
 RL Submitted (MAY-2001) to the EMBL/GenBank/DBJ databases.

CC -1- SIMILARITY: CONTAINS 1 RING-TYPE ZINC FINGER.  
 DR EMBL; AK027687; BAB5297.1; -  
 DR InterPro; IPR000253; FHA.  
 DR InterPro; IPR008984; SMAD FHA.  
 DR InterPro; IPR001841; Znf\_ring.  
 DR Pfam; PF00498; FHA; 1.  
 DR Pfam; PF00097; zf-C3HC4; 1.  
 DR SMART; SM00240; FHA; 1.  
 DR SMART; SM00184; RING; 1.  
 DR PROSITE; PS5006; FHA DOMAIN; 1.  
 DR PROSITE; PS00518; ZF\_RING\_1; 1.  
 DR PROSITE; PS50089; ZF\_RING\_2; 1.  
 KM Hypothetical protein: Metal-binding; Zinc; Zinc-finger.  
 SQ SEQUENCE 652 AA; 72030 MW; 572F2CE6D1743D80 CRC64;

Query Match 7.4%; Score 89; DB 4; Length 652;  
 Best Local Similarity 23.0%; Pred. No. 6.8; Indels 74; Gaps 11;  
 Matches 52; Conservative 26; Mismatches 74;

QY 8 HRSRCATPPRGDFC-----GGTERAIDQA-----SFTSMEMDTQVVGSSPL----- 50  
 Db 422 YRQQAQPPH--CPAEGEGAPQALGDAPSTSVSLTTAVQDVYVCPLOGSHALCTCCFQ 478  
 QY 51 ---GPAGLGAEPAPAGPOLPSWLOPERCAVFOCAQCHAVLADSVHLAMDLSR----- 99  
 Db 479 PMPDRAREEDPPRVAPQ-----QCAVC---LQPFCHLYWGCTRTGCGYGLA 522  
 QY 100 -----SLGAVVFSRYTNVNLVLEAPFLVIGESLKGSTY-NILFCGSCGIPVGFHLVSTHA 153  
 Db 523 PFCENLGDGKCLDGVLNNSYESDILKNYLAT-RGLTWKNNL-----TES 566

QY 154 ALAALRGHFCLS-----SDKWVCYLTKTAIVNASEMDIQNVPLSE 194  
 Db 567 LVALQGVFLSDRYRTGTVLYCYCGGLRSFRELTYQYKQNI PAS 612

RESULT 10  
 Q9NR74 PRELIMINARY; PRT; 664 AA.

AC Q9NR74; 01-OCT-2000 (TREMBLrel. 15, Created)  
 DT 01-OCT-2000 (TREMBLrel. 15, last sequence update)  
 DT 01-OCT-2003 (TREMBLrel. 25, last annotation update)  
 DE Cell cycle checkpoint protein CHFR.  
 OS Homo sapiens (human).  
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
 OC Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RX MEDLINE=20388685; PubMed=10935642;  
 RA Scolnick D.W., Halazonetis T.D.,  
 RT "Chfr defines a mitotic stress checkpoint that delays entry into  
 RT metaphase";  
 RL Nature 406:430-435(2000).  
 CC -1- SIMILARITY: CONTAINS 1 RING-TYPE ZINC FINGER.  
 DR EMBL; ARI70724; AAR91084.1; -  
 DR PDB; 1LGP; 07-AUG-02.  
 DR Genew; HGNC:20455; CHFR.  
 DR GO; GO:0007093; P:mitotic checkpoint; TAS.  
 DR InterPro; IPR000253; FHA.  
 DR InterPro; IPR008984; SMAD FHA.  
 DR InterPro; IPR001841; Znf\_ring.  
 DR Pfam; PF00498; FHA; 1.  
 DR Pfam; PF00097; zf-C3HC4; 1.  
 DR SMART; SM00240; FHA; 1.  
 DR SMART; SM00184; RING; 1.  
 DR PROSITE; PS5006; FHA DOMAIN; 1.  
 DR PROSITE; PS00518; ZF\_RING\_1; 1.  
 DR PROSITE; PS50089; ZF\_RING\_2; 1.  
 KM Metal-binding; Zinc; Zinc-finger.  
 SQ SEQUENCE 664 AA; 73386 MW; 141A1E7FEFA36A2 CRC64;

Query Match 7.4%; Score 89; DB 4; Length 664;  
 Best Local Similarity 23.0%; Pred. No. 6.9; Indels 74; Gaps 11;  
 Matches 52; Conservative 26; Mismatches 74;

QY 8 HRSRCATPPRGDFC-----GGTERAIDQA-----SFTSMEMDTQVVGSSPL----- 50  
 Db 434 YRQQAQPPH--CPAEGEGAPQALGDAPSTSVSLTTAVQDVYVCPLOGSHALCTCCFQ 490  
 QY 51 ---GPAGLGAEPAPAGPOLPSWLOPERCAVFOCAQCHAVLADSVHLAMDLSR----- 99  
 Db 491 PMPDRAREEDPPRVAPQ-----QCAVC---LQPFCHLYWGCTRTGCGYGLA 534  
 QY 100 -----SLGAVVFSRYTNVNLVLEAPFLVIGESLKGSTY-NILFCGSCGIPVGFHLVSTHA 153  
 Db 535 PFCENLGDGKCLDGVLNNSYESDILKNYLAT-RGLTWKNNL-----TES 578

QY 154 ALAALRGHFCLS-----SDKWVCYLTKTAIVNASEMDIQNVPLSE 194  
 Db 579 LVALQGVFLSDRYRTGTVLYCYCGGLRSFRELTYQYKQNI PAS 624

RESULT 11  
 Q8BUZ9 PRELIMINARY; PRT; 663 AA.

AC Q8BUZ9; 01-MAR-2003 (TREMBLrel. 23, Created)  
 DT 01-MAR-2003 (TREMBLrel. 23, last sequence update)  
 DT 01-OCT-2003 (TREMBLrel. 25, last annotation update)  
 DE Similar to cell cycle checkpoint protein CHFR.  
 GN 5730484M2ORIK.  
 OS Mus musculus (Mouse).  
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
 OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RC STRAIN=C57BL/6J; TISSUE=Body;  
 RX MEDLINE=22354683; PubMed=12466851;  
 RA The FANTOM Consortium,  
 RA The RIKEN Genome Exploration Research Group Phase I & II Team;  
 RT "Analysis of the mouse transcriptome based on functional annotation of  
 RT 60,770 full-length cDNAs";  
 RL Nature 420:563-573 (2002).  
 DR EMBL; AK077629; BAC36912.1; -  
 DR MGD; MGI:2444898; 5730484M2ORIK.  
 DR InterPro; IPR000253; FHA.  
 DR InterPro; IPR008984; SMAD FHA.  
 DR InterPro; IPR001841; Znf\_ring.  
 DR Pfam; PF00498; FHA; 1.  
 DR Pfam; PF00097; zf-C3HC4; 1.  
 DR SMART; SM00240; FHA; 1.  
 DR SMART; SM00184; RING; 2.  
 DR PROSITE; PS5006; FHA DOMAIN; 1.  
 DR PROSITE; PS00518; ZF\_RING\_1; 1.  
 DR PROSITE; PS50089; ZF\_RING\_2; 1.  
 SQ SEQUENCE 663 AA; 73855 MW; 42BD653DC84ADP45 CRC64;

Query Match 7.3%; Score 87.5; DB 11; Length 663;  
 Best Local Similarity 23.3%; Pred. No. 9.7;  
 Matches 49; Conservative 19; Mismatches 71; Indels 71; Gaps 9;

QY 38 EMDTQVVGK-----SSPLCPA-GLGAEAPAPAGPOLPS-----WLOP- 72  
 Db 432 EYRQAVOSLPCVPESELGATLALGGEAPSTASLPTADYWCPLQSHALCTCCFQPM 491  
 QY 73 -----ERCAVFOCAQCHAVLADSVHLAMDLSR-----SLGAVVFSRV 109  
 Db 492 PDRAREEDPPRVAPQCAVC---LQPFCHLYWGCTRTGCGYGLA 548  
 QY 110 TNNVLEAPFLVIGESLKGSTY-NILFCGSCGIPVGFHLVSTHAALALRGHFCLS----- 165  
 Db 549 LNNNSYESDILKNYLAT-RGLTWKSVL-----TESLALQGVFLMSDYRI 593

```

QY      166 -SDRMVCYLLTKTKAIVNASEMDIONVPLSE 194
      : : : : : : : : : : : : : : : : : :
Db      594 TGNVLCYCCGGRSFRRELTYYRONIPASE 623

RESULT 12
Q810L3 PRELIMINARY; PRT; 664 AA.
ID Q810L3
AC Q810L3
DT 01-JUN-2003 (Tremblrel. 24, Created)
DT 01-JUN-2003 (Tremblrel. 24, Last sequence update)
DT 01-OCT-2003 (Tremblrel. 25, Last annotation update)
DE RIKEN cDNA 5730484M20 gene.
OS Mus musculus (Mouse).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
OX NCBI_TaxId=10090;
RN [1]
RP SEQUENCE FROM N.A.
RC TISSUE=Limb;
RA Strausberg R.;
RL Submitted (MAR-2003) to the EMBL/GenBank/DBJ databases.
DR EMBL; BC049792; AAH49792.1; -
DR InterPro; IPR000253; FHA.
DR InterPro; IPR008984; SMAD_FHA.
DR InterPro; IPR001841; Znf_Ring.
DR Pfam; PF00498; FHA; 1.
DR Pfam; PF00097; zf-C3HC4; 1.
DR SMART; SM00240; FHA; 1.
DR SMART; SM00184; RING; 2.
DR PROSITE; PSS0006; FHA_DOMAIN; 1.
DR PROSITE; PS00518; ZF_RING_1; 1.
DR PROSITE; PS50089; ZF_RING_2; 1.
SQ SEQUENCE 664 AA; 73871 MW; D615B3E3F463DEB6 CRC64;

Query Match
Best Local Similarity 7.3%; Score 87; DB 11; Length 664;
Matches 46; Conservative 20; Mismatches 76; Indels 66; Gaps 8;

QY      38 EMDTQVVKG-----SSPLGPA-GIGAEPNAGPOLPSMLQPERC-----AVFGA--- 81
      : : : : : : : : : : : : : : : : : :
Db      432 EYRRQAVQGLPCPVPSESLGATILAGGAPSTASAPTAAPYMCPLQSHAICTCCFP 491

QY      82 -----QCHAVLADSVHLAWDLR-----SLGAVFSRVN 111
      : : : : : : : : : : : : : : : : : :
Db      492 MPDRRAERQDPVVAQQCAVCLQPCCHLMGCTRGCGCIAPFCELMLGDKCIDGVN 551

QY      112 NVVLAPPLVIGESLKGSTVNLFCGSGCIPVGFHLVSTHAALARGHFLCS-----S 166
      : : : : : : : : : : : : : : : : : :
Db      552 NNVSDDILKNYLAT-RGLTWKRSVL-----TESILALQGVFWLSDYRITG 596

QY      167 DKMVCYLLTKTKAIVNASEMDIONVPLSE 194
      : : : : : : : : : : : : : : : : : :
Db      597 NTVLCCGGRSFRRELTYYRONIPASE 624

RESULT 13
Q97FQ5 PRELIMINARY; PRT; 669 AA.
ID Q97FQ5
AC Q97FQ5
DT 01-OCT-2001 (Tremblrel. 18, Created)
DT 01-OCT-2001 (Tremblrel. 18, Last sequence update)
DT 01-OCT-2003 (Tremblrel. 25, Last annotation update)
DE MAD-dependent DNA ligase.
OS CAC2673.
OC Clostridium acetobutylicum.
OC Bacteria; Firmicutes; Clostridia; Clostridiales; Clostridiaceae;
OC Clostridium.
OX NCBI_TaxId=1488;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=ATCC 824 / DSM 792 / VKM B-1787;

```

```

RX MEDLINE=21359325; PubMed=11466286;
RA Noelling J., Breton G., Omelchenko M.V., Makarova K.S., Zeng Q.,
RA Gibson R., Lee H.M., Dubois J., Qiu D., Hitti J., Wolf Y.I.,
RA Tatusov R.L., Sabathe F., Doucette-Stamm L., Soucaille P., Daly M.J.,
RA Bennett G.N., Koonin E.V., Smith D.R.;
RT "Genome sequence and comparative analysis of the solvent-producing
RT bacterium Clostridium acetobutylicum.";
RL J. Bacteriol. 183:4823-4838(2001).
DR EMBL; AE007765; AAK80620.1; -.
DR PIR; A97229; A97229.
DR GO; GO:0005622; C:intracellular; IEA.
DR GO; GO:0016020; C:membrane; IEA.
DR GO; GO:0003677; F:DNA binding; IEA.
DR GO; GO:0003911; F:DNA binding (NAD) activity; IEA.
DR GO; GO:0016874; F:ligase activity; IEA.
DR GO; GO:0005215; F:transporter activity; IEA.
DR GO; GO:0006281; P:DNA repair; IEA.
DR GO; GO:0006260; P:DNA replication; IEA.
DR GO; GO:0006810; P:transport; IEA.
DR InterPro; IPR001357; BRCT.
DR InterPro; IPR001679; DNALigase.
DR InterPro; IPR004150; DNA_ligase_OB.
DR InterPro; IPR000445; HhH.
DR InterPro; IPR003583; HhH.
DR InterPro; IPR008984; NucLeic_acid_OB.
DR InterPro; IPR005829; Sug_transporter.
DR InterPro; IPR004149; Znf_DNALigase_C4.
DR Pfam; PF00533; BRCT; 1.
DR Pfam; PF01653; DNA_ligase_N; 1.
DR Pfam; PF03120; DNA_ligase_OB; 1.
DR Pfam; PF03119; DNA_ligase_ZBD; 1.
DR Pfam; PF00633; HhH; 2.
DR ProDom; PD003944; DNALigase; 1.
DR SMART; SM00292; BRCT; 1.
DR SMART; SM00278; HhH; 3.
DR SMART; SM00532; LIGANC; 1.
DR TIGRPFAM; TIGR00575; dnj; 1.
DR PROSITE; PSS0172; BRCT; 1.
DR PROSITE; PS01056; DNA_LIGASE_N2; 1.
DR PROSITE; PS00216; SUGAR_TRANSPORT_1; 1.
SQ SEQUENCE 669 AA; 75900 MW; A4A4C9CAB72767DF CRC64;

Query Match
Best Local Similarity 7.2%; Score 86.5; DB 16; Length 669;
Matches 37; Conservative 22; Mismatches 45; Indels 31; Gaps 8;

QY      96 DLNR-----SLGAVFERVNNVLEAPFLVIGESLKGSTVNL-----LFGSGCGIPVG 146
      : : : : : : : : : : : : : : : : : :
Db      366 DGRKKVKIGSRVFRKSNVYI--PEIMGVTEETEGETNEIEAPTCPCGSEIVKEGV 422

QY      147 HLVSTHAALARGHFLCSDRMVCYLLTKTKAIVNASEMDIONVPLSEKIAE-LKEKIV 204
      : : : : : : : : : : : : : : : : : :
Db      423 HL-----FC--EMTLSCPKQMWYSIYHFAREAMNIGFSEKTAELQLEK-- 465

QY      205 LTHNRUKSLMKILSE 219
      : : : : : : : : : : : : : : : : : :
Db      466 LNIKSIDLVYRITKE 480

RESULT 14
O839T4 PRELIMINARY; PRT; 145 AA.
ID O839T4
AC O839T4
DT 01-JUN-2003 (Tremblrel. 24, Created)
DT 01-JUN-2003 (Tremblrel. 24, Last sequence update)
DT 01-JUN-2003 (Tremblrel. 24, Last annotation update)
DE Hypothetical protein.
OS EF0068.
OC Enterococcus faecalis (Streptococcus faecalis).
OC Bacteria; Firmicutes; Lactobacillales; Enterococcaceae; Enterococcus.
OX NCBI_TaxId=1351;
RN [1]

```

```

RP SEQUENCE FROM N.A.
RC STRAIN=V583 / ATCC 700802;
RX MEDLINE=22550857; PubMed=12663927;
RA Paulsen I.T., Banerjee L., Myers G.S.A., Nelson K.E., Seshadri R.,
RA Read T.D., Pouts D.E., Eisen J.A., Gill S.R., Heidelberg J.F.,
RA Tetteijn H., Dodson R.J., Umayam L., Brinkac L., Beaman M.,
RA Daugherty S., Deboy R.T., Durkin S., Kolonay J., Madupu R., Nelson W.,
RA Vamathevan J., Tran B., Upton J., Hansen T., Shetty J., Khouri H.,
RA Ullrichback T., Radue D., Ketchum K.A., Dougherty B.A., Fraser C.M.,
RT "Role of mobile DNA in the evolution of vancomycin-resistant
  Enterococcus faecalis.";
RL Science 299:2071-2074(2003).
DR EMBL; AE016947; AA079948.1; -.
DR TIGR; EF0068; -.
SQ SEQUENCE 145 AA; 16506 MW; 81A8BD4A7F8CE17 CRC64;

Query Match 7.2%; Score 86; DB 16; Length 145;
Best Local Similarity 26.5%; Pred. No. 1.9;
Matches 40; Conservative 27; Mismatches 54; Indels 30; Gaps 9;

QY 87 LADSVHAMDLSRSIGAVV--FSRYTNNTVLEA--PLVIGESLKG-----STYN 133
DB 4 LREYIYLA SEQPLSTGSIITLNGLSKPTGLINKAFEPFINDPEGDPPEKFSYSNYLKTIN 63
QY 134 LIFGSCGIPVGFHYSTHAAAL--RGHFCLSDDKWVCYLTKTAIVNASMDIONV 190
DB 64 YQVCGMCSLPNDQNIETPSALALGYIRNH-----KATC---TTKYLIALANGPEKK 115
QY 191 PLSE---KIAELKEKIVLTNNRLKSLMKIL 217
DB 116 PISKVYTRKLSLSYK-DLYQEMK-LTKII 144

```

## RESULT 15

```

Q82H01 PRELIMINARY; PRT; 337 AA.
AC Q82H01;
DT 01-JUN-2003 (TREMBlrel. 24, Created)
DT 01-JUN-2003 (TREMBlrel. 24, Last sequence update)
DT 01-OCT-2003 (TREMBlrel. 25, Last annotation update)
DE Putative tryptophanyl-tRNA synthetase.
GN TRPS1 OR SAV3417.
OS Streptomyces avermitilis.
OC Bacteria; Actinobacteria; Actinomycetales;
OC Streptomycineae; Streptomycetaceae; Streptomyces.
OX NCBI_TaxID=33903;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=MA-4680 / ATCC 31267 / NCIMB 12804 / NRRL 8165;
RX MEDLINE=21477403; PubMed=11572948;
RA Omura S., Ikeda H., Ishikawa J., Hanamoto A., Takahashi C.,
RA Shinose M., Takahashi Y., Horikawa H., Nakazawa H., Osone T.,
RA Kikuchi H., Shiba T., Sakaki Y., Hattori M.;
RT "Genome sequence of an industrial microorganism Streptomyces
  avermitilis: deducing the ability of producing secondary
  metabolites.";
RT Proc. Natl. Acad. Sci. U.S.A. 98:12215-12220(2001).
RN [2]
RP SEQUENCE FROM N.A.
RC STRAIN=MA-4680 / ATCC 31267 / NCIMB 12804 / NRRL 8165;
RX MEDLINE=22608306; PubMed=12692562;
RA Ikeda H., Ishikawa J., Hanamoto A., Shinose M., Kikuchi H., Shiba T.,
RA Sakaki Y., Hattori M., Omura S.;
RT "Complete genome sequence and comparative analysis of the industrial
  microorganism Streptomyces avermitilis.";
RT Nat. Biotechnol. 21:526-531(2003).
RL EMBL; AP05035; BAC71129.1; -.
DR GO; GO:0005524; F:ATP binding; IEA.
DR GO; GO:0004830; F:tryptophan-tRNA ligase activity; IEA.
DR GO; GO:0006436; P:tryptophanyl-tRNA aminoacylation; IEA.
DR InterPro; IPR002305; trna-synt_1b.
DR InterPro; IPR001412; trna-synt_1.

```

```

DR InterPro; IPR002306; Trp trna-synt_1b.
DR Pfam; PF00579; trna-synt_1b; 1.
DR PRINTS; PR01039; TRNASYNTHTRP.
DR TIGRFAMs; TIGR00233; trps; 1.
DR PROSITE; PS00178; AA trna ligase I; 1.
KW Aminoacyl-tRNA synthetase; Complete proteome.
SQ SEQUENCE 337 AA; 37096 MW; 23DAF4D121131864 CRC64;

Query Match 7.1%; Score 85; DB 16; Length 337;
Best Local Similarity 24.7%; Pred. No. 7.1;
Matches 55; Conservative 34; Mismatches 100; Indels 34; Gaps 11;

QY 18 GDFGCGTER--AIDQASFTSMEDTQVKGSSPLGAGIAGEPPAAGPQ--LPSWLOER 74
DB 22 GNYLGAVRQWVALQESHDAFTMYVDLHAI--TVQDPADIDANTRLAAQILLAGIDBER 79
QY 75 CAVFQCAQCHAVLADSVHAMDLSRSIGAVVFSRYTNNTVLEAFPVIGESLKG----- 130
DB 80 CTLF--VQSH--VEHAGLAWIMNCTLFGCSASHMTQFKDSAK-----QADRASVGLF 130
QY 131 TY-----NLIFGSCGIPVG---FHYSTHAAALRGHFCLSDDKWVCYLTKTAIV 180
DB 131 TYPVLQVADILLYQANEVVPVEDRQHLIELTRDLAERFNGRFGETFTVPKYYILKETAKI 190
QY 181 NASEMDIQ--NVPLEKIAELKEKIVLTNNRLKSLMKILSEVT 221
DB 191 ---FDLQDPPIKMSKASSTPKGLINLDEPKATKKVSAVT 229

```

Search completed: July 20, 2004, 11:17:44  
Job time : 42 secs



GenCore version 5.1.6  
Copyright (c) 1993 - 2004 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: July 20, 2004, 11:10:48 ; Search time 13 Seconds  
(without alignments)

917.236 Million cell updates/sec

Title: US-09-942-052a-728

Perfect score: 1198  
Sequence: 1 MAOPLRHRSCATPPRGDF.....LKSIMKILSEVTPDQSKPEN 229

Scoring table: BLOSUM62  
Gapop 10.0 , Gapext 0.5

Searched: 141681 seqs, 52070155 residues

Total number of hits satisfying chosen parameters: 141681

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%  
Maximum Match 100%

Listing first 45 summaries

Database : SwissProt\_42.\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

#### SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	1198	100.0	229	1	OTF5_HUMAN
2	113	9.4	233	1	CU45_HUMAN
3	105	8.8	155	1	YC2C_SCHPO
4	89	7.4	893	1	POL2_BAMMA
5	88	7.3	250	1	KXK7_CAMJE
6	85	7.1	337	1	SYW1_STRAW
7	85	7.1	2194	1	GLSN_MEDSA
8	82.5	6.9	891	1	POL2_BAMN
9	81.5	6.8	457	1	UIDB_ECOLI
10	81.5	6.8	640	1	UI06_HSV7
11	81.5	6.8	662	1	UI06_HSV6
12	81.5	6.8	3119	1	CA1C_MOUSE
13	81	6.8	371	1	YN06_SYNEL
14	81	6.8	894	1	POL2_BAMW
15	81	6.8	1418	1	CE11_CABEL
16	80.5	6.7	1711	1	PTPV_RAT
17	80	6.7	469	1	IDG2_YEAST
18	79	6.6	308	1	KDGD_OCEIH
19	79	6.6	1705	1	PTPV_MOUSE
20	78.5	6.6	404	1	PTTB_ARATH
21	78.5	6.6	697	1	ULAA_HCMVA
22	78	6.5	351	1	LPXD_VIBCH
23	77.5	6.5	485	1	GATA_CLOTE
24	77.5	6.5	662	1	UI06_HSV6
25	77	6.4	738	1	TREX_HUMAN
26	76.5	6.4	307	1	GPDA_ANASP
27	76	6.3	639	1	EPG_RHIME
28	75.5	6.3	310	1	MCAL_MOUSE
29	75.5	6.3	700	1	UVRD_MYCTU
30	75.5	6.3	888	1	SYA_CORGL
31	75.5	6.3	1534	1	DNM1_ARATH
32	75.5	6.3	1723	1	LY75_MOUSE
33	75	6.3	1010	1	CLPP_CHLEU

34	75	6.3	1374	1	YO3D_SCHPO	009884 schizosach
35	74.5	6.2	1057	1	RAG1_BRARE	013033 brachydanio
36	74	6.2	282	1	END4_AQUAE	067551 aquifex ae
37	74	6.2	367	1	GLD4_ECOLI	P32655 escherichia
38	74	6.2	377	1	GBB_ARATH	P49177 arabidopsis
39	74	6.2	463	1	NHR1_CABEL	Q9xt14 caenorhabdi
40	74	6.2	621	1	NIOL_BUCHEP	089at6 buchera ap
41	74	6.2	2109	1	RRLP_VSVSU	P03523 vesicular s
42	74	6.2	3063	1	CA1C_HUMAN	Q99715 homo sapien
43	73.5	6.1	401	1	NH65_CABEL	Q45907 caenorhabdi
44	73.5	6.1	403	1	MHP1_ECOLI	P75899 escherichia
45	73.5	6.1	1015	1	DNL3_MOUSE	P97386 mus musculu

#### ALIGNMENTS

RESULT 1  
ID OTF5\_HUMAN STANDARD; PRT; 229 AA.  
AC 043482; Q96BX7;  
DT 28-FEB-2003 (Rel. 41, Created)  
DT 28-FEB-2003 (Rel. 41, Last sequence update)  
DT 10-OCT-2003 (Rel. 42, Last annotation update)  
DE Opa-interacting protein 5.  
GN OTF5.  
OS Homo sapiens (Human).  
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
OC Mammalia; Eutheria; Primates; Catarrhini; Hominiidae; Homo.  
OX NCBI\_TaxID=9606;  
RN [1]  
RP SEQUENCE FROM N.A.  
RX MEDLINE=98125741; PubMed=9466265;  
RA Williams J.M., Chen G.-C., Zhu L., Rest R.F.;  
RT "Using the Yeast two-hybrid system to identify human epithelial cell  
RT proteins that bind gonococcal Opa proteins: intracellular gonococci  
RT bind pyruvate kinase via their Opa proteins and require host pyruvate  
RT for growth.";  
RL Mol. Microbiol. 27:171-186(1998).  
RN [2]  
RP SEQUENCE FROM N.A.  
RC TISUE=Uterus;  
RX MEDLINE=22388257; PubMed=12477932;  
RA Strausberg R.L., Feingold E.A., Grouse L.H., Derge J.G.,  
RA Klausner R.D., Collins F.S., Wagner L., Shennan C.M., Schuler G.D.,  
RA Altschul S.F., Zeeberg B., Buetow K.H., Schaefer C.F., Bat N.K.,  
RA Hopkins R.F., Jordan H., Moore T., Max S.I., Wang J., Hsieh F.,  
RA Diatchenko L., Marnettina K., Farmer A.A., Rubin G.M., Hong L.,  
RA Stapleton M., Soares M.B., Bonaldo M.F., Casavant T.L., Scheetz T.E.,  
RA Brownstein M.J., Ustin T.B., Toshnyuk S., Carninci P., Prange C.,  
RA Raha S.S., Loquellano N.A., Peters G.J., Abramson R.D., Mullany S.J.,  
RA Bosak S.A., McEwan P.J., McKernan K.J., Malek J.A., Gunaratne P.H.,  
RA Richards S., Worley K.C., Hale S., Garcia A.M., Gay L.J., Hulyk S.W.,  
RA Villalón D.K., Muzny D.M., Sodergren E.J., Lu X., Gibbs R.A.,  
RA Fahey J., Helton E., Kettelman M., Madan A., Shevchenko Y., Bouffard G.G.,  
RA Blakesley R.W., Touchman J.W., Green E.D., Dixon M.C.,  
RA Rodriguez A.C., Grimwood J., Schmutz J., Myers R.M.,  
RA Buterfield Y.S.N., Krzywinski M.I., Skalska U., Smalhus D.E.,  
RA Schermer A., Schein J.E., Jones S.J.M., Marra M.A.;  
RT "Generation and initial analysis of more than 15,000 full-length  
RT human and mouse cDNA sequences.";  
RL Proc. Natl. Acad. Sci. U.S.A. 99:16899-16903(2002).  
CC -!- SUBUNIT: Binds outer membrane protein OpaF from Neisseria  
CC gonorrhoeae.  
CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
CC the European Bioinformatics Institute. There are no restrictions on its  
CC use by non-profit institutions as long as its content is in no way  
CC modified and this statement is not removed. Usage by and for commercial  
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).

Query Match	Best Local Similarity	Score	DB 1	Length	229
Matches 229; Conservative	100.0%; Pred. No. 2,7e-104;	0	Mismatches	0	Indels
		0	Gaps	0	

  

Query	1	MAAPLRRHSRCATPRGDFCCGTERAIDQASFTTSMEMDVTVKGSPLGAGI	GAEP	60
Db	1	MAAPLRRHSRCATPRGDFCCGTERAIDQASFTTSMEMDVTVKGSPLGAGI <td>GAEP</td> <td>60</td>	GAEP	60
Qy	61	AAAGPOLPSWLOPERCAVFCQACHVLADSVYLANDLSRSI	GAVFESRVTNNVYLEAPFL	120
Db	61	AAAGPOLPSWLOPERCAVFCQACHVLADSVYLANDLSRSI	GAVFESRVTNNVYLEAPFL	120
Qy	121	VGIESGLKSTYVNLFCGSCGIPVGFHLYSTHAPALALRGHFC	LSDDMVCYLKTKATV	180
Db	121	VGIESGLKSTYVNLFCGSCGIPVGFHLYSTHAPALALRGHFC	LSDDMVCYLKTKATV	180
Qy	181	NASEMDIQVPLSEKIAELKEKIVITNHRKLSMKILISVTPDQSKPEN	229	
Db	181	NASEMDIQVPLSEKIAELKEKIVITNHRKLSMKILISVTPDQSKPEN	229	

  

RESULT 2	ID	CU45 HUMAN	STANDARD:	PRT:	233 AA.
AC	Q9NVE9;				
DT	16-OCT-2001 (Rel. 40, Created)				
DT	16-OCT-2001 (Rel. 40, Last sequence update)				
DT	10-OCT-2003 (Rel. 42, Last annotation update)				
DE	Purative protein C21orf45.				
GN	C21ORF45.				
OS	Homo sapiens (Human).				
OC	Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi;				
OC	Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.				
OX	NCBI_Taxid=9606;				
RN	[1]				
RP	SEQUENCE FROM N.A.				
EX	MEDLINE=20237674; PubMed=10773462;				
RA	Slavov D., Hattori M., Sakaki Y., Rosenthal A., Shimizu N.,				
RA	Mimoshima S., Kudoh J., Yaspo M.L., Ramser J., Reinhardt R.,				
RA	Reimer C., Clancy K., Ryndtich A., Gardner K.;				
RT	"Criteria for gene identification and features of genome organization:				
RT	analysis of 6.5 Mb of DNA sequence from human chromosome 21.";				
RL	Gene 247:215-232(2000).				
RN	[2]				
RP	SEQUENCE FROM N.A.				
RC	TISSUE=Blood;				
XX	MEDLINE=22388257; PubMed=12477932;				
RA	Strausberg R.L., Peingold E.A., Grouse L.H., Derge J.G.,				
RA	Klauser R.D., Collins F.S., Wagner L., Shenmen C.M., Schuler G.D.,				
RA	Altschul S.F., Zeeberg B., Buetow K.H., Schaefer C.F., Bhat N.K.,				
RA	Hopkins R.F., Jordan K., Moore T., Max S.I., Wang J., Hsieh F.,				
RA	Diatchenko L., Marsina K., Farmer A.A., Rubin G.M., Hong L.,				
RA	Stapleton M., Soares M.B., Bonaldo M.F., Casavant T.L., Scheetz T.E.,				
RA	Brownstein M.J., Udwin T.B., Toshilyuk S., Cammelli P., Prange C.,				
RA	Rada S.S., Loquellano N.A., Peters G.J., Abramson R.D., Mullaly S.J.,				
RA	Richards S.A., McGwan P.D., McKernan K.J., Malek J.A., Gamaralle P.H.,				
RA	Richards S., Worley K.C., Hale S., Garcia A.M., Gay L.J., Hulyk S.W.,				
RA	Vallalon D.K., Muzny D.M., Sodergren E.J., Lu X., Gibbs R.A.,				
RA	Faney J., Helton E., Kettelman M., Madan A., Rodriguez S., Sanchez A.,				
RA	Whitting M., Madan A., Young A.C., Shevchenko Y., Bouffard G.G.,				
RA	Blakesley R.W., Touchman J.W., Green E.D., Dickson M.C.,				
RA	Rodriguez A.C., Grimwood J., Schmutz J., Myers R.M.,				
RA	Butterfield Y.S.N., Krzywinski M.I., Skalska U., Smalins D.E.,				
RA	Schmerch A., Schein J.E., Jones S.J.W., Maria M.A.,				
RT	"Generation and initial analysis of more than 15,000 full-length				

human ad mouse cDNA sequences." RT  
Proc. Natl. Acad. Sci. U.S.A. 97:16899-16903 (2002).  
-1- SIMILARITY: TO S.POMBE C970.12.

-----  
CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
CC between the Swiss Institute of Bioinformatics and the EMBL outstation  
CC at the European Bioinformatics Institute. There are no restrictions on its  
CC use by non-profit institutions as long as its content is in no way  
CC modified and this statement is not removed. Usage by and for commercial  
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).

CC  
DR EMBL, AF231921; AAF72945.1; -  
DR EMBL, BC042917; AAH42917.1; -  
DR Genew; HGNC:1286; C21orf45.  
KW Hypothetical protein.  
SQ SEQUENCE 233 AA; 25863 MW; A07522806CAB6221 CRC64;

Query Match 9.4%; Score 113; DB 1; Length 233;  
Best Local Similarity 27.1%; Pred. No. 0.0032;  
Matches 52; Conservative 30; Mismatches 88; Indels 22; Gaps 8;

DQ 31 ASFTTSEMHTQVYKGGSPPLGPAAGLGAEEPAAGPQLPSWLQPERCAVFOCAQHAVLADS 90  
Db 49 ASWMSMSSEDAV-----ADVERAQL-EEENAAAE-----ERPLVFLSCGCRRLGDS 95

DQ 91 VHLAMDLSR-SLGAVFPSRTYNNVLEAPFLVIGISLKGSTNNLIFCGSCGIPVGFHL 149  
Db 96 --LSWVASQEDTNCILIRCVSCVSVYDKKIKLSREKENGCVLETLCCAGCSINTLGYVR 153

DQ 150 STHAALALRGHPCLSSDKMYCYLL--KTQAYVASEMDIONVPLSEKIAELKEIKVLTH 207  
Db 154 CTPEKMLDYKKRDLFCLSVEALIESYVLGSSSEKQIV-SEDKELFNL---ESRVEIEKLTQWE 209

DQ 208 NRLKSLMKTIIE 219  
Db 210 DVLKALQMKLME 221

-----

RESULT 3  
YCZC SCHPO  
ID YCZC SCHPO STANDARD; PRT; 155 AA.  
AC 09F802;  
DT 16-OCT-2001 (Rel. 40, Created)  
DT 16-OCT-2001 (Rel. 40, Last sequence update)  
DT 28-FEB-2003 (Rel. 41, Last annotation update)  
DE Hypothetical protein C970.12 in chromosome III.  
GN SPC970.12.  
OS Schizosaccharomyces pombe (fission yeast).  
OC Eukaryota; Fungi; Ascomycota; Schizosaccharomycetes;  
OC Schizosaccharomycetales; Schizosaccharomycetaceae;  
OC Schizosaccharomycetes.  
OX NCBI\_TaxID=4896;  
RN RN  
RP  
SEQUENCE FROM N.A.  
STRAIN=972;  
MEDLINE=21848401; PubMed=11859360;  
RX Wood V., Gwilliam R., Rajandream M.A., Lyne M., Lyne R., Stewart A.,  
Sgourou V., Peat N., Hayles J., Baker S., Basham D., Bowman S.,  
Brooks K., Brown D., Brown S., Chillingworth T., Churcher C.M.,  
Collins M., Connor R., Cronin A., Davis P., Feltwell T., Fraser A.,  
Gentles S., Goble A., Howarth S., Harris D., Hidalgo J., Hodgson G.,  
James K., Jones L., Jones M., Leather S., McDonald S., McLean J.,  
Mooney P., Moule S., Mungall K., Murphy L., Niblett D., Odeli C.,  
Oliver K., O'Neill S., Pearson D., Quail M.A., Rabinowitch E.,  
Rutherford K., Rutter S., Saunders D., Seeger K., Sharp S.,  
Skellon J., Simmonds M., Squares R., Squares S., Stevens K.,  
Taylor K., Taylor R.G., Tivey A., Walsh S.V., Warren T., Whitehead S.,  
Woodward J., Volckaert G., Aert R., Kobben J., Grymopre B.,  
Welters I., Vanstreels E., Rieger M., Schaefer M., Mueller-Auer S.,  
Gabel C., Fuchs M., Fritz C., Holzer E., Moestl D., Hilbert H.,  
Bozrym K., Langer I., Beck A., Lehrach H., Reinhardt R., Pohl T.M.

```

RA Eger P., Zimmermann W., Wedler H., Wambutt R., Purnelle B.,
RA Goffeau A., Cadieu E., Dreano S., Gloux S., Lelaine V., Mottier S.,
RA Gilbert F., Aves S.J., Xiang Z., Hunt C., Moore K., Hust S.M.,
RA Lucas M., Roher M., Gallardin C., Tallada V.A., Garzon A., Rhode G.,
RA Daga R.R., Cruzado L., Jimenez J., Sanchez M., del Rey F., Benito J.,
RA Dominguez A., Revuelta J.L., Moreno S., Armstrong J., Forsburg S.L.,
RA Cerutti L., Lowe T., McCombie M.R., Paulsen I., Potashkin J.,
RA Shpakovski G.V., Ussery D., Barrell B.G., Nurse P.,
RT "The genome sequence of Schizosaccharomyces pombe."
RL Nature 415:871-880(2002).
CC -1- SIMILARITY: TO HUMAN C21ORF45.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
DR EMBL: AL031530; CAB72327.1; -.
DR GeneDB: SPombe; SPCC970.12; -.
KM Hypothetical protein.
SQ SEQUENCE 155 AA: 17874 MW: 17874 MW: A7AEDDF93760B0B CRC64:

Query Match      8.8%; Score 105; DB 1; Length 155;
Best Local Similarity 30.1%; Pred. No. 0.011;
Matches 41; Conservative 18; Mismatches 61; Indels 16; Gaps 6;

OY 71 QPERCAVFOCAQCHAVLADSVHLAMDLSRS--LGAVFSRYTNNVLEAPFLVGEGLK 128
DB 19 QP---SVFOCKKCFQIVGDS--NANVISHREYLSFTLSDAVENSRYVBDTEFRSDG-- 71
OY 129 GSTVLLFCGSGGIVGFPHLYSTHAALALRGHFLCSDKVCYLLTKKAIYNASEMDIQ 188
DB 72 -CVSELSCTRNCEVIGKRVNSTPIYLDIRMTYFSDKQAYLGKNT-VNPEGLTRY 129
OY 189 NVPLSEKIAELKEKIV 204
DB 130 QVLD-----EMREDII 140

RESULT 4
POL2 BAMVA
ID POL2 BAMVA STANDARD; PRT; 893 AA.
AC Q65329;
DT 15-DEC-1998 (Rel. 37, Created)
DT 15-DEC-1998 (Rel. 37, Last sequence update)
DT 28-FEB-2003 (Rel. 41, Last annotation update)
DE Genome polyprotein 2 (Contains: Helper component proteinase
DE (EC 3.4.22.45) (HC-Pro); 70 kDa protein).
OS Barley mild mosaic virus (Strain ASL) (BaMV).
OC Viruses; ssRNA positive-strand viruses, no DNA stage; Polyviridae;
OC Bymovirus.
OX NCBI_TaxID=103899;
RN [1]
RP SEQUENCE FROM N.A.
RA "Time U", Kuehne T.;
RT "The complete nucleotide sequence of RNA2 of barley mild mosaic virus
RT (BaMV)."
RL Eur. J. Plant Pathol. 100:233-241(1994).
CC -1- CATALYTIC ACTIVITY: Hydrolyzes a Gly-|-Gly bond at its own C-
CC terminus, commonly in the sequence -Tyr-Xaa-Val-Gly-|-Gly, in the
CC processing of the polyviral polyprotein.
CC -1- SIMILARITY: HC PROTEINASE BELONGS TO PEPTIDASE FAMILY C6.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).

```

```

CC EMBL: X75933; CA53537.1; -.
DR MEROPS; C06.001; -.
DR InterPro; IPR001337; TMV coat.
DR Pfam; PF00721; TMV coat; 1.
KW Polypeptide; Hydrolase; Thiol protease.
FT CHAIN 1 229 HELPER COMPONENT PROTEINASE (POTENTIAL).
FT CHAIN 230 893 70 KDA PROTEIN.
FT ACT SITE 117 117 POTENTIAL.
FT ACT SITE 189 189 POTENTIAL.
FT SITE 229 230 CLEAVAGE (BY HC-PRO PROTEASE)
FT SITE (POTENTIAL).
SQ SEQUENCE 893 AA: 98328 MW: D6B74D9D6C6407 CRC64:

Query Match      7.4%; Score 89; DB 1; Length 893;
Best Local Similarity 22.3%; Pred. No. 3;
Matches 79; Conservative 36; Mismatches 105; Indels 134; Gaps 18;

OY 4 QPLHRSGKATPPGCD--FCGGERALDQ-ASPTTSMW-----DTQVKGSSPLGPA 53
DB 13 QVLR---RFSIPTSGDRILVNSTDQPLGFGARFDTSLQTLSTQVTDPEVLKQKSN-PT 68
OY 54 GLG-AEPPAGPOLPSW-----LQPERCAVFOCAQCHAVL----- 87
DB 69 HDVAVYLEASRFRFPWFELTNSCTRGSGTHAQNLQAFATAEKSGFCYNNLLPLSFD 128
OY 88 ----ADSVH-LAMDLSRLGCAVVSRYTNNVLEA---PFLV----- 121
DB 129 IIDAHDSFRGFVQPLDITLGAVPSLSVNLVMMHAAVRFELIASPIPTIAFAESLQF 188
OY 122 -----GLEGS--LKG-STYNLLFCGSGGIVGFPHLYSTHA-----LAALRGHFLCS 166
DB 189 HVTDRKGVPMGNITLKACRVYELLADAGIGCEYMLPVGAAPQYSFWKKSMDHF--TS 246
OY 167 DKMVCYLTKTAIVYASEMD-----IQN----- 189
DB 247 DRFVFLMQLLSALAEQDVAITHARDALLSALONAGYTNVVAHERFPNGHDSIWL 306
OY 190 ----VPLSEKIAELKEKIV-----LTHNRLSKMLKILSEVTPDOOSREN 229
DB 307 NLSEAPISEKLTETLKRVLVGHRSDDTADITHNVQHVEVLKTMVSQFSKTTN 360

RESULT 5
KKA7 CAMOE
ID KKA7 CAMOE STANDARD; PRT; 250 AA.
AC P14508;
DT 01-JAN-1990 (Rel. 13, Created)
DT 01-JAN-1990 (Rel. 13, Last sequence update)
DT 28-FEB-2003 (Rel. 41, Last annotation update)
DE Aminoglycoside 3'-phosphotransferase (EC 2.7.1.95) (Kanamycin kinase,
DE type VII) (Neomycin-kanamycin phosphotransferase, type VII)
DE (Aph(3')VII).
GN APHA-7.
OS Campylobacter jejuni.
OC Bacteria; Proteobacteria; Epsilonproteobacteria; Campylobacteriales;
OC Campylobacteraceae; Campylobacter.
OX NCBI_TaxID=197;
RN [1]
RP SEQUENCE FROM N.A.
RA STRAIN=PS11178;
RC MEDLINE=89387451; PubMed=2550983;
RX Tenover F.C., Gilbert T., O'Hara P.;
RA "Nucleotide sequence of a novel kanamycin resistance gene, aphA-7,
RT from Campylobacter jejuni and comparison to other kanamycin
RT phosphotransferase genes."
RL Plasmid 22:52-58(1989).
CC -1- FUNCTION: RESISTANCE TO KANAMYCIN AND STRUCTURALLY RELATED
CC AMINOGLYCOSIDES, INCLUDING AMIKACIN.
CC -1- CATALYTIC ACTIVITY: ATP + kanamycin = ADP + kanamycin 3'-
CC phosphate.
CC -1- SIMILARITY: TO OTHER AMINOGLYCOSIDE PHOSPHOTRANSFERASES.

```

-----  
CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
CC the European Bioinformatics Institute. There are no restrictions on its  
CC use by non-profit institutions as long as its content is in no way  
CC modified and this statement is not removed. Usage by and for commercial  
CC entities requires a license agreement (See <http://www.ebi.ac.uk/announcements/>  
CC or send an email to [license@ebi.ac.uk](mailto:license@ebi.ac.uk)).

DR EMBL; M29953; AAA76822.1; -  
DR PIR; A43623; A43623.  
DR InterPro; IPR002575; APH.  
DR Pfam; PF01636; APH; 1.  
KW Antibiotic resistance; Transferase; Kinase; ATP-binding; Plasmid.  
FT ACT SITE 178 BY SIMILARITY.  
SQ SEQUENCE 250 AA; 29688 MW; AF3B7ADBE85533BC CRC64;

Query Match 7.3%; Score 88; DB 1; Length 250;  
Best Local Similarity 28.7%; Pred. No. 0.74;  
Matches 33; Conservative 18; Mismatches 46; Indels 18; Gaps 5;

QY 105 VFSRVNNTNVLEAPFVIGESLKGSTYNLFCGSGIPVGFHYSTHAALALRG-HFC 163  
Db 40 IFEKITYSVKREBEMMMMLSDKLKVPDV-----IEYVREHSEYILMSLRGKHID 90  
QY 164 LSSDKVVCYLKTKATVNA-----SENDIQNVPLSEKI-AELKEKIVLTHNRKSL 213  
Db 91 CFIDHPIKTY---ECLVNALHQLQALDINCPSSKIDVRLKELKYLIDNRIRADI 142

RESULT 6  
STYL STRAW STANDARD; PRT; 337 AA.  
ID SYW1 STRAW  
AC 082H01;  
DT 15-MAR-2004 (Rel. 43, Last sequence update)  
DT 15-MAR-2004 (Rel. 43, Last annotation update)  
DE 1) (Trypan 1).  
DE 1) (Trypan 1).  
GN TRPS1 OR SAV3417.  
OS Streptomyces avermitilis.  
OC Bacteria; Actinobacteriae; Actinobacteridae; Actinomycetales;  
OC Streptomycinae; Streptomycetaceae; Streptomycetes.  
OX NCBI\_TaxID=33903;  
RN [1]  
RP SEQUENCE FROM N.A.  
RC STRAIN-MA-4680 / ATCC 31267 / NCIMB 12804 / NRRL 8165;  
RX MEDLINE=21477403; PubMed=11572948;  
RA Omura S., Ikeda H., Ishikawa J., Hanamoto A., Takahashi C.,  
RA Kikuchi H., Shibata T., Horikawa H., Nakazawa H., Osone T.,  
RA "Genome sequence of an industrial microorganism Streptomyces  
RT avermitilis: deducing the ability of producing secondary  
RT metabolites".  
RT Proc. Natl. Acad. Sci. U.S.A. 98:12215-12220 (2001).  
RN [2]  
RP SEQUENCE FROM N.A.  
RC STRAIN-MA-4680 / ATCC 31267 / NCIMB 12804 / NRRL 8165;  
RX MEDLINE=22608306; PubMed=12692562;  
RA Ikeda H., Ishikawa J., Hanamoto A., Shinose M., Kikuchi H., Shiba T.,  
RA Sasaki Y., Hattori M., Omura S.;  
RT "Complete genome sequence and comparative analysis of the industrial  
RT microorganism Streptomyces avermitilis".  
RL Nat. Biotechnol. 21:526-531 (2003).  
CC -1- CATALYTIC ACTIVITY: ATP + L-tryptophan + tRNA(Trp) = AMP +  
CC diphosphate + L-tryptophanyl-tRNA(Trp).  
CC -1- SUBUNIT: Homodimer (By similarity).  
CC -1- SUBCELLULAR LOCATION: Cytoplasmic.  
CC -1- SIMILARITY: Belongs to class-I aminoacyl-tRNA synthetase family.  
CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
CC the European Bioinformatics Institute. There are no restrictions on its

-----  
CC use by non-profit institutions as long as its content is in no way  
CC modified and this statement is not removed. Usage by and for commercial  
CC entities requires a license agreement (See <http://www.ebi.ac.uk/announcements/>  
CC or send an email to [license@ebi.ac.uk](mailto:license@ebi.ac.uk)).

DR EMBL; AP005035; BAC71129.1; -  
DR HAMAP; MF\_00140; -; 1.  
DR InterPro; IPR002305; tRNA-synt\_1b.  
DR InterPro; IPR001412; tRNA-synt\_1.  
DR Pfam; PF00579; tRNA-synt\_1b; 1.  
DR PRINTS; PR01039; TRNASYNTHTRP.  
DR TIGRPFAM; TIGR00233; tRPS; 1.  
DR PROSITE; PS00178; AA tRNA LIGASE 1; 1.  
KW Aminoacyl-tRNA synthetase; Protein biosynthesis; Ligase; ATP-binding;  
KW Complete proteome.  
FT SITE 14 23 "HIGH" REGION.  
FT SITE 199 203 "KMSK" REGION.  
FT BINDING 202 202 ATP (By SIMILARITY).  
SQ SEQUENCE 337 AA; 37096 MW; 23DAF4D121131864 CRC64;

Query Match 7.1%; Score 85; DB 1; Length 337;  
Best Local Similarity 24.7%; Pred. No. 2.1;  
Matches 55; Conservative 34; Mismatches 100; Indels 34; Gaps 11;

QY 18 GDFCGGTER--AIDQASFTTSMEDTQVVGKSGSPGAGAEPAAGPQ-IPSWLOPER 74  
Db 22 GNYLGAIRQWVALIOESHAFAFYVVDLHAI--TVGPQPADLRANTRILAAQLLAAGLDPER 79  
QY 75 CAVFQAOCHAVLAQSVHLAMDLNRSLSGAVNVSRYNNVLEAPFVIGESLKG----- 130  
Db 80 CULF--VQSH--VPEHAQLAMTNCITGFGASRMTQFKDKSKA-----QADRAASVGLF 130  
QY 131 TY-----NLIFCGSGCGLPVG-----FHLVSTAAALALRGHFLCSDKMVCYLKTKATV 180  
Db 131 TVPVLQVADILLYQANENVVGEDQHQHILTTDLAERNGRGERTVTPKPIILAKETAKI 190  
QY 181 NASEMDIQ--NVPLSEKIAELKEKIVLTHNRKSLMKILSEVT 221  
Db 191 ----FDLPDSIKMSKASSTPKGLINLIDEPATAKVKSATV 229

RESULT 7  
GLSN MEDSA  
ID GLSN\_MEDSA STANDARD; PRT; 2194 AA.  
AC Q03460;  
DT 01-OCT-1993 (Rel. 27, Created)  
DT 01-OCT-1993 (Rel. 27, Last sequence update)  
DT 28-FEB-2003 (Rel. 41, Last annotation update)  
DE Glutamate synthase [NADH], chloroplast precursor (EC 1.4.1.14) (NADH-  
DE GOGAT).  
OS Medicago sativa (Alfalfa).  
OC Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;  
OC Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots; rosids;  
OC eustosids I; Fabales; Fabaceae; Papilionoideae; Trifoliaceae; Medicago.  
OX NCBI\_TaxID=3879;  
RN [1]  
RP SEQUENCE FROM N.A., AND SEQUENCE OF 102-114.  
RX MEDLINE=93200806; PubMed=8453303;  
RA Gregerson R.G., Miller S.S., Twary S.N., Gantt J.S., Vance C.P.;  
RT "Molecular characterization of NADH-dependent glutamate synthase from  
RT alfalfa nodules".  
RL Plant Cell 5:215-226 (1993).  
CC -1- CATALYTIC ACTIVITY: 2 L-glutamate + NAD(+) = L-glutamine + 2-  
CC oxoglutarate + NADH.  
CC -1- COFACTOR: Binds a 3Fe-4S cluster; FAD and FMN.  
CC -1- PATHWAY: Glutamine synthetase/GOGAT pathway which is involved in  
CC the assimilation of ammonia.  
CC -1- SUBUNIT: Monomer.  
CC -1- SUBCELLULAR LOCATION: AMYLOPLAST (POTENTIAL).  
CC -1- TISSUE SPECIFICITY: Root nodules.  
CC -1- SIMILARITY: TO OTHER GLUTAMATE SYNTHASIS.

CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
CC the European Bioinformatics Institute. There are no restrictions on its  
CC use by non-profit institutions as long as its content is in no way  
CC modified and this statement is not removed. Usage by and for commercial  
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).

DR EMBL: L01660; ABA4617.1; -  
DR PIR: J01977; J01977.  
DR InterPro: IPR000755; Adnrx\_reductase.  
DR InterPro: IPR002489; DUF14.  
DR InterPro: IPR001327; FAD\_pyr\_redox.  
DR InterPro: IPR006982; Glu synth centr.  
DR InterPro: IPR006981; Glu synth centr.  
DR InterPro: IPR002932; Glu synthase.  
DR InterPro: IPR006005; Glu synth sub1.  
DR InterPro: IPR001100; Pyr\_redox.  
DR InterPro: IPR001003; Pyridine\_redox\_2.  
DR Pfam: PF04898; Glu syn central; 1.  
DR Pfam: PF04897; Glu synth NTN; 1.  
DR Pfam: PF01645; Glu synthase; 1.  
DR Pfam: PF01493; GSKG; 1.  
DR Pfam: PF00070; Pyr\_redox; 1.  
DR PRINTS: PR00419; ADXRDPASE.  
DR PRINTS: PR00368; FADPDR.  
DR PRINTS: PR00411; PNDRTASE1.  
DR PRINTS: PR00469; PNDRTASE1.  
DR TIGRPMAS: TIGR01317; GOGAT sm. gam; 1.  
DR Oxidoreductase; Iron-sulfur; 3Fe-4S; Flavoprotein; FAD; FMN; NAD;  
KW Chloroplast; Amyloplast; Transil peptide; Glutamate biosynthesis.  
FT TRANSIT 1 101  
FT CHAIN 102 2194 GLUTAMATE SYNTHASE [NADH].  
FT DOMAIN 102 456 GLUTAMATE AMIDOTRANSFERASE (POTENTIAL).  
FT NP\_BIND 1193 1250 FMN (BY SIMILARITY).  
FT METAL 1246 1246 IRON-SULFUR (3FE-4S) (BY SIMILARITY).  
FT METAL 1252 1252 IRON-SULFUR (3FE-4S) (BY SIMILARITY).  
FT METAL 1257 1257 IRON-SULFUR (3FE-4S) (BY SIMILARITY).  
FT NP\_BIND 1974 1988 NAD (POTENTIAL).  
SQ SEQUENCE 2194 AA; 240373 MW; 370A1B0F178367C4 CRC64;

Query Match 7.1%; Score 85; DB 1; Length 2194;  
Best Local Similarity 24.7%; Pred. No. 22;  
Matches 58; Conservative 22; Mismatches 79; Indels 76; Gaps 12;

QY 4 QPIRRRCATPPRRDPCGTERAIDQAS-----FTTSMEMDQVVGSSPLDPA 53  
DB 1037 EPLADGSR--NEKRS-----AIKQVASGRFGVSYLLTNADLQIKMAQGAAP----- 1082  
QY 54 GLGAEPP-----AAGPOLPSWLOPERCAVFOCAQCHAVLADSVHLMADLSR 99  
DB 1083 GEGGLPGHKVIGDAITRNSAVGILIS--PRPHNDIYS-----IEDLAQLHDLKN 1133  
QY 100 SLGAVFVRVTNNVLEAPFVIGFGLSKGTYNLLFCG-----SCGIPYVG 145  
DB 1134 ANPA---ARISVYKVSAGVGIAGVGVGHAHEVILSGHDGTGASRWGTGKISAGLPEW 1190  
QY 146 FHLYSTHALLA--LRGHCLSSDKMVCYLKLT-----KATVNSEMDIQVPL 192  
DB 1191 LGLAETHQTLVANDLRGRTTLQTDGQ---LKTGRDVAIAALLGAEVGFSTAPL 1241

## RESULT 8

PO12\_BAMN STANDARD; PRT; 891 AA.  
ID POL2\_BAMN  
AC P89684;  
DT 15-DEC-1998 (Rel. 37, Created)  
DT 15-DEC-1998 (Rel. 37, Last sequence update)  
DT 28-FEB-2003 (Rel. 41, Last annotation update)  
DE Genome polyprotein 2 [Contains: Helper component proteinase  
DE (EC 3.4.22.45) (HC-Pro); 70 kDa protein].  
OS Barley mild mosaic virus (strain Nal) (BamMV).  
OC Viruses; ssRNA positive-strand viruses, no DNA stage; Polyviridae;

CC Bymovirus.  
OX NCBI\_TaxID=103900;  
RN [1]  
SEQUENCE FROM N.A.  
RX MEDLINE=97128984; Pubmed=8973524;  
RA Kashwazaki S.;  
RT "The complete nucleotide sequence and genome organization of barley  
RT mild mosaic virus (Nal strain).";  
RL Arch. Virol. 141:2077-2083(1996).  
CC -1- CATALYTIC ACTIVITY: Hydrolyzes a Gly-Gly bond at its own C-  
CC terminus, commonly in the sequence -Tyr-Xaa-Val-Gly-Gly- in the  
CC processing of the polyviral polyprotein.  
CC -1- SMILARITY: HC PROTEINASE BELONGS TO PEPTIDASE FAMILY C6.

CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
CC the European Bioinformatics Institute. There are no restrictions on its  
CC use by non-profit institutions as long as its content is in no way  
CC modified and this statement is not removed. Usage by and for commercial  
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).

DR EMBL: D83409; BAA18954.1; -  
DR InterPro: IPR001337; TMV\_coat.  
DR Pfam: PF00721; TMV\_coat; 1.  
KW Polyprotein; Hydrolyase; Thiol protease.  
FT CHAIN 1 229 HELPER COMPONENT PROTEINASE (POTENTIAL).  
FT ACT\_SITE 230 891 70 kDa PROTEIN.  
FT ACT\_SITE 117 117 POTENTIAL.  
FT ACT\_SITE 189 189 POTENTIAL.  
FT SITE 229 230 CLEAVAGE (BY HC-PRO PROTEASE)  
FT SITE 229 230 (POTENTIAL).  
SQ SEQUENCE 891 AA; 97942 MW; A290247196822BE2 CRC64;

Query Match 6.9%; Score 82.5; DB 1; Length 891;  
Best Local Similarity 21.3%; Pred. No. 12;  
Matches 74; Conservative 38; Mismatches 104; Indels 111; Gaps 17;

QY 11 RCATPPRRD---FCGTERAIDQ-ASFTTSMEW-----DQVVGSSPLDPAIG-ABE 59  
DB 17 RFSIPASGRLLISNPTDQPIGLFCGAFDTSIQVGDDEPVKOKIHI-PTHLDISA 75  
QY 60 PAAGPOLBSW-----LQPERCAVFOCAQCHAVL-----AD 89  
DB 76 LEASRSPFWLFTNSFCTFGSGJHAQNLQAPATAEFGSGCYNNLLVPLSPDITDARD 135  
QY 90 SVHL-AMDLRSLSLGAVFVRVTNNVLEA---PFLV-----G 122  
DB 136 SFRVVEQDLPMLGKAPSLSVLVNMLAATRFPEIVASPVPTIAFDAESLQFHTDKRG 195  
QY 123 IEGS---LK-GSTYNLLFCGSGCIPVGFHLYSTHAA-----LAALRGHCLSSDKMVCYL 173  
DB 196 VPGMNNILKAGRYELLSLADGVGCEYMLYVGAAPQYSPWKKSMDF--TSDFVEFL 253  
QY 174 LKTKAIVASEMD-----ION-----VPL 192  
DB 254 AMQNLASALBEDYTTTHDALALALQAGTYNNVAERPPNCHDSITWLNISEAPI 313  
QY 193 SEKIAELKEKIV-----LTHNRKLSLMLKILSEVTPDQSKPEN 229  
DB 314 SEKLDLKRKYLIVGHRSDTADITNHHQYVEVAKTMSVQSKSTN 360

## RESULT 9

UIDB\_ECOLI STANDARD; PRT; 457 AA.  
ID UIDB\_ECOLI  
AC P30868; P77457;  
DT 01-JUL-1993 (Rel. 26, Created)  
DT 01-NOV-1997 (Rel. 35, Last sequence update)  
DT 16-OCT-2001 (Rel. 40, Last annotation update)  
DE Glucuronide carrier protein (Glucuronide permease).  
GN uidB OR GUSB OR uidP OR B1616.  
OS Escherichia coli.

OC Bacteria; Proteobacteria; Gammaproteobacteria; Enterobacteriales;  
 OC Enterobacteriaceae; Escherichia.  
 OX NCBI\_TaxID=562;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RA Jefferson R.A.;  
 RL Submitted (OCT-1994) to the EMBL/GenBank/DBJ databases.  
 RN [2]  
 RP SEQUENCE FROM N.A.  
 RC STRAIN=K12 / MG1655;  
 RX MEDLINE=9742617; PubMed=9278503;  
 RA Blattner F.R., Plunkett G. III, Bloch C.A., Perna N.T., Burland V.,  
 RA Riley M., Colado-Vides J., Glasner J.D., Rode C.K., Mayhew G.F.,  
 RA Gregor J., Davis N.W., Kirkpatrick H.A., Goeden M.A., Rose D.J.,  
 RA Mau B., Shao Y.;  
 RL "The complete genome sequence of Escherichia coli K-12.";  
 RL Science 277:1453-1474(1997).  
 RN [3]  
 RP SEQUENCE FROM N.A.  
 RC STRAIN=K12;  
 RX MEDLINE=97251357; PubMed=9097039;  
 RA Alba H., Baba T., Fujita K., Hayashi K., Inada T., Isono K.,  
 RA Itoh T., Kasai H., Kashimoto K., Kimura S., Kitakawa M.,  
 RA Kitagawa M., Makino K., Miki T., Mizobuchi K., Mori H., Mori T.,  
 RA Motomura K., Nakade S., Nakamura Y., Nishimoto H., Nishio Y.,  
 RA Oshima T., Saito N., Sampei G., Seki Y., Sivasubram S.,  
 RA Tagami H., Takeda J., Takemoto K., Takeuchi Y., Wada C.,  
 RA Yamamoto Y., Horiiuchi T.;  
 RL "A 570-kb DNA sequence of the Escherichia coli K-12 genome  
 corresponding to the 28.0-40.1 min region on the linkage map.";  
 RL DNA Res. 3:363-377(1996).  
 RN [4]  
 RP PRELIMINARY SEQUENCE OF 1-112 FROM N.A.  
 RX MEDLINE=87041472; PubMed=3534890;  
 RA Jefferson R.A., Burgess S.M., Hirsch D.;  
 RL "Beta-glucuronidase from Escherichia coli as a gene-fusion marker.";  
 RL Proc. Natl. Acad. Sci. U.S.A. 83:8447-8451(1986).  
 CC -!- SUBCELLULAR LOCATION: Integral membrane protein. Inner membrane  
 CC (potential).  
 CC -!- SIMILARITY: BELONGS TO THE SODIUM:GALACTOSIDE SYMPORTER FAMILY  
 CC (SGF).  
 CC -----  
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
 CC the European Bioinformatics Institute. There are no restrictions on its  
 CC use by non-profit institutions as long as its content is in no way  
 CC modified and this statement is not removed. Usage by and for commercial  
 CC entities requires a license agreement (see http://www.isb-sib.ch/announce/  
 CC or send an email to license@isb-sib.ch).  
 CC -----  
 DR EMBL; M14641; AAA68924.1; -;  
 DR EMBL; AE000257; AAC74688.1; -;  
 DR EMBL; D90805; BAA15367.1; -;  
 DR PIR; B64918; B64918.  
 DR ECGene; EG11658; uidB.  
 DR InterPro; IPR001927; Na/Gal\_symport.  
 DR TIGRfams; TIGR00792; gph; 1.  
 DR PROSITE; PS00872; NA\_GALACTOSIDE\_SYMPT, 1.  
 KW Transport; Sugar transport; Transmembrane; Inner membrane; Symport;  
 KW Complete proteome.  
 KM  
 CC  
 FT TRANSMEM 22 42 POTENTIAL.  
 FT TRANSMEM 44 64 POTENTIAL.  
 FT TRANSMEM 80 100 POTENTIAL.  
 FT TRANSMEM 109 129 POTENTIAL.  
 FT TRANSMEM 147 167 POTENTIAL.  
 FT TRANSMEM 181 201 POTENTIAL.  
 FT TRANSMEM 229 249 POTENTIAL.  
 FT TRANSMEM 264 284 POTENTIAL.  
 FT TRANSMEM 309 329 POTENTIAL.  
 FT TRANSMEM 373 393 POTENTIAL.  
 FT TRANSMEM 409 429 POTENTIAL.  
 FT TRANSMEM 457 477 POTENTIAL.  
 FT TRANSMEM 499 519 POTENTIAL.  
 FT TRANSMEM 547 567 POTENTIAL.  
 FT TRANSMEM 599 619 POTENTIAL.  
 FT TRANSMEM 647 667 POTENTIAL.  
 FT TRANSMEM 699 719 POTENTIAL.  
 FT TRANSMEM 747 767 POTENTIAL.  
 FT TRANSMEM 799 819 POTENTIAL.  
 FT TRANSMEM 847 867 POTENTIAL.  
 FT TRANSMEM 899 919 POTENTIAL.  
 FT TRANSMEM 947 967 POTENTIAL.  
 FT TRANSMEM 999 1019 POTENTIAL.  
 FT TRANSMEM 1047 1067 POTENTIAL.  
 FT TRANSMEM 1099 1119 POTENTIAL.  
 FT TRANSMEM 1147 1167 POTENTIAL.  
 FT TRANSMEM 1199 1219 POTENTIAL.  
 FT TRANSMEM 1247 1267 POTENTIAL.  
 FT TRANSMEM 1299 1319 POTENTIAL.  
 FT TRANSMEM 1347 1367 POTENTIAL.  
 FT TRANSMEM 1399 1419 POTENTIAL.  
 FT TRANSMEM 1447 1467 POTENTIAL.  
 FT TRANSMEM 1499 1519 POTENTIAL.  
 FT TRANSMEM 1547 1567 POTENTIAL.  
 FT TRANSMEM 1599 1619 POTENTIAL.  
 FT TRANSMEM 1647 1667 POTENTIAL.  
 FT TRANSMEM 1699 1719 POTENTIAL.  
 FT TRANSMEM 1747 1767 POTENTIAL.  
 FT TRANSMEM 1799 1819 POTENTIAL.  
 FT TRANSMEM 1847 1867 POTENTIAL.  
 FT TRANSMEM 1899 1919 POTENTIAL.  
 FT TRANSMEM 1947 1967 POTENTIAL.  
 FT TRANSMEM 1999 2019 POTENTIAL.  
 FT TRANSMEM 2047 2067 POTENTIAL.  
 FT TRANSMEM 2099 2119 POTENTIAL.  
 FT TRANSMEM 2147 2167 POTENTIAL.  
 FT TRANSMEM 2199 2219 POTENTIAL.  
 FT TRANSMEM 2247 2267 POTENTIAL.  
 FT TRANSMEM 2299 2319 POTENTIAL.  
 FT TRANSMEM 2347 2367 POTENTIAL.  
 FT TRANSMEM 2399 2419 POTENTIAL.  
 FT TRANSMEM 2447 2467 POTENTIAL.  
 FT TRANSMEM 2499 2519 POTENTIAL.  
 FT TRANSMEM 2547 2567 POTENTIAL.  
 FT TRANSMEM 2599 2619 POTENTIAL.  
 FT TRANSMEM 2647 2667 POTENTIAL.  
 FT TRANSMEM 2699 2719 POTENTIAL.  
 FT TRANSMEM 2747 2767 POTENTIAL.  
 FT TRANSMEM 2799 2819 POTENTIAL.  
 FT TRANSMEM 2847 2867 POTENTIAL.  
 FT TRANSMEM 2899 2919 POTENTIAL.  
 FT TRANSMEM 2947 2967 POTENTIAL.  
 FT TRANSMEM 2999 3019 POTENTIAL.  
 FT TRANSMEM 3047 3067 POTENTIAL.  
 FT TRANSMEM 3099 3119 POTENTIAL.  
 FT TRANSMEM 3147 3167 POTENTIAL.  
 FT TRANSMEM 3199 3219 POTENTIAL.  
 FT TRANSMEM 3247 3267 POTENTIAL.  
 FT TRANSMEM 3299 3319 POTENTIAL.  
 FT TRANSMEM 3347 3367 POTENTIAL.  
 FT TRANSMEM 3399 3419 POTENTIAL.  
 FT TRANSMEM 3447 3467 POTENTIAL.  
 FT TRANSMEM 3499 3519 POTENTIAL.  
 FT TRANSMEM 3547 3567 POTENTIAL.  
 FT TRANSMEM 3599 3619 POTENTIAL.  
 FT TRANSMEM 3647 3667 POTENTIAL.  
 FT TRANSMEM 3699 3719 POTENTIAL.  
 FT TRANSMEM 3747 3767 POTENTIAL.  
 FT TRANSMEM 3799 3819 POTENTIAL.  
 FT TRANSMEM 3847 3867 POTENTIAL.  
 FT TRANSMEM 3899 3919 POTENTIAL.  
 FT TRANSMEM 3947 3967 POTENTIAL.  
 FT TRANSMEM 3999 4019 POTENTIAL.  
 FT TRANSMEM 4047 4067 POTENTIAL.  
 FT TRANSMEM 4099 4119 POTENTIAL.  
 FT TRANSMEM 4147 4167 POTENTIAL.  
 FT TRANSMEM 4199 4219 POTENTIAL.  
 FT TRANSMEM 4247 4267 POTENTIAL.  
 FT TRANSMEM 4299 4319 POTENTIAL.  
 FT TRANSMEM 4347 4367 POTENTIAL.  
 FT TRANSMEM 4399 4419 POTENTIAL.  
 FT TRANSMEM 4447 4467 POTENTIAL.  
 FT TRANSMEM 4499 4519 POTENTIAL.  
 FT TRANSMEM 4547 4567 POTENTIAL.  
 FT TRANSMEM 4599 4619 POTENTIAL.  
 FT TRANSMEM 4647 4667 POTENTIAL.  
 FT TRANSMEM 4699 4719 POTENTIAL.  
 FT TRANSMEM 4747 4767 POTENTIAL.  
 FT TRANSMEM 4799 4819 POTENTIAL.  
 FT TRANSMEM 4847 4867 POTENTIAL.  
 FT TRANSMEM 4899 4919 POTENTIAL.  
 FT TRANSMEM 4947 4967 POTENTIAL.  
 FT TRANSMEM 4999 5019 POTENTIAL.  
 FT TRANSMEM 5047 5067 POTENTIAL.  
 FT TRANSMEM 5099 5119 POTENTIAL.  
 FT TRANSMEM 5147 5167 POTENTIAL.  
 FT TRANSMEM 5199 5219 POTENTIAL.  
 FT TRANSMEM 5247 5267 POTENTIAL.  
 FT TRANSMEM 5299 5319 POTENTIAL.  
 FT TRANSMEM 5347 5367 POTENTIAL.  
 FT TRANSMEM 5399 5419 POTENTIAL.  
 FT TRANSMEM 5447 5467 POTENTIAL.  
 FT TRANSMEM 5499 5519 POTENTIAL.  
 FT TRANSMEM 5547 5567 POTENTIAL.  
 FT TRANSMEM 5599 5619 POTENTIAL.  
 FT TRANSMEM 5647 5667 POTENTIAL.  
 FT TRANSMEM 5699 5719 POTENTIAL.  
 FT TRANSMEM 5747 5767 POTENTIAL.  
 FT TRANSMEM 5799 5819 POTENTIAL.  
 FT TRANSMEM 5847 5867 POTENTIAL.  
 FT TRANSMEM 5899 5919 POTENTIAL.  
 FT TRANSMEM 5947 5967 POTENTIAL.  
 FT TRANSMEM 5999 6019 POTENTIAL.  
 FT TRANSMEM 6047 6067 POTENTIAL.  
 FT TRANSMEM 6099 6119 POTENTIAL.  
 FT TRANSMEM 6147 6167 POTENTIAL.  
 FT TRANSMEM 6199 6219 POTENTIAL.  
 FT TRANSMEM 6247 6267 POTENTIAL.  
 FT TRANSMEM 6299 6319 POTENTIAL.  
 FT TRANSMEM 6347 6367 POTENTIAL.  
 FT TRANSMEM 6399 6419 POTENTIAL.  
 FT TRANSMEM 6447 6467 POTENTIAL.  
 FT TRANSMEM 6499 6519 POTENTIAL.  
 FT TRANSMEM 6547 6567 POTENTIAL.  
 FT TRANSMEM 6599 6619 POTENTIAL.  
 FT TRANSMEM 6647 6667 POTENTIAL.  
 FT TRANSMEM 6699 6719 POTENTIAL.  
 FT TRANSMEM 6747 6767 POTENTIAL.  
 FT TRANSMEM 6799 6819 POTENTIAL.  
 FT TRANSMEM 6847 6867 POTENTIAL.  
 FT TRANSMEM 6899 6919 POTENTIAL.  
 FT TRANSMEM 6947 6967 POTENTIAL.  
 FT TRANSMEM 6999 7019 POTENTIAL.  
 FT TRANSMEM 7047 7067 POTENTIAL.  
 FT TRANSMEM 7099 7119 POTENTIAL.  
 FT TRANSMEM 7147 7167 POTENTIAL.  
 FT TRANSMEM 7199 7219 POTENTIAL.  
 FT TRANSMEM 7247 7267 POTENTIAL.  
 FT TRANSMEM 7299 7319 POTENTIAL.  
 FT TRANSMEM 7347 7367 POTENTIAL.  
 FT TRANSMEM 7399 7419 POTENTIAL.  
 FT TRANSMEM 7447 7467 POTENTIAL.  
 FT TRANSMEM 7499 7519 POTENTIAL.  
 FT TRANSMEM 7547 7567 POTENTIAL.  
 FT TRANSMEM 7599 7619 POTENTIAL.  
 FT TRANSMEM 7647 7667 POTENTIAL.  
 FT TRANSMEM 7699 7719 POTENTIAL.  
 FT TRANSMEM 7747 7767 POTENTIAL.  
 FT TRANSMEM 7799 7819 POTENTIAL.  
 FT TRANSMEM 7847 7867 POTENTIAL.  
 FT TRANSMEM 7899 7919 POTENTIAL.  
 FT TRANSMEM 7947 7967 POTENTIAL.  
 FT TRANSMEM 7999 8019 POTENTIAL.  
 FT TRANSMEM 8047 8067 POTENTIAL.  
 FT TRANSMEM 8099 8119 POTENTIAL.  
 FT TRANSMEM 8147 8167 POTENTIAL.  
 FT TRANSMEM 8199 8219 POTENTIAL.  
 FT TRANSMEM 8247 8267 POTENTIAL.  
 FT TRANSMEM 8299 8319 POTENTIAL.  
 FT TRANSMEM 8347 8367 POTENTIAL.  
 FT TRANSMEM 8399 8419 POTENTIAL.  
 FT TRANSMEM 8447 8467 POTENTIAL.  
 FT TRANSMEM 8499 8519 POTENTIAL.  
 FT TRANSMEM 8547 8567 POTENTIAL.  
 FT TRANSMEM 8599 8619 POTENTIAL.  
 FT TRANSMEM 8647 8667 POTENTIAL.  
 FT TRANSMEM 8699 8719 POTENTIAL.  
 FT TRANSMEM 8747 8767 POTENTIAL.  
 FT TRANSMEM 8799 8819 POTENTIAL.  
 FT TRANSMEM 8847 8867 POTENTIAL.  
 FT TRANSMEM 8899 8919 POTENTIAL.  
 FT TRANSMEM 8947 8967 POTENTIAL.  
 FT TRANSMEM 8999 9019 POTENTIAL.  
 FT TRANSMEM 9047 9067 POTENTIAL.  
 FT TRANSMEM 9099 9119 POTENTIAL.  
 FT TRANSMEM 9147 9167 POTENTIAL.  
 FT TRANSMEM 9199 9219 POTENTIAL.  
 FT TRANSMEM 9247 9267 POTENTIAL.  
 FT TRANSMEM 9299 9319 POTENTIAL.  
 FT TRANSMEM 9347 9367 POTENTIAL.  
 FT TRANSMEM 9399 9419 POTENTIAL.  
 FT TRANSMEM 9447 9467 POTENTIAL.  
 FT TRANSMEM 9499 9519 POTENTIAL.  
 FT TRANSMEM 9547 9567 POTENTIAL.  
 FT TRANSMEM 9599 9619 POTENTIAL.  
 FT TRANSMEM 9647 9667 POTENTIAL.  
 FT TRANSMEM 9699 9719 POTENTIAL.  
 FT TRANSMEM 9747 9767 POTENTIAL.  
 FT TRANSMEM 9799 9819 POTENTIAL.  
 FT TRANSMEM 9847 9867 POTENTIAL.  
 FT TRANSMEM 9899 9919 POTENTIAL.  
 FT TRANSMEM 9947 9967 POTENTIAL.  
 FT TRANSMEM 9999 10019 POTENTIAL.  
 FT TRANSMEM 10047 10067 POTENTIAL.  
 FT TRANSMEM 10099 10119 POTENTIAL.  
 FT TRANSMEM 10147 10167 POTENTIAL.  
 FT TRANSMEM 10199 10219 POTENTIAL.  
 FT TRANSMEM 10247 10267 POTENTIAL.  
 FT TRANSMEM 10299 10319 POTENTIAL.  
 FT TRANSMEM 10347 10367 POTENTIAL.  
 FT TRANSMEM 10399 10419 POTENTIAL.  
 FT TRANSMEM 10447 10467 POTENTIAL.  
 FT TRANSMEM 10499 10519 POTENTIAL.  
 FT TRANSMEM 10547 10567 POTENTIAL.  
 FT TRANSMEM 10599 10619 POTENTIAL.  
 FT TRANSMEM 10647 10667 POTENTIAL.  
 FT TRANSMEM 10699 10719 POTENTIAL.  
 FT TRANSMEM 10747 10767 POTENTIAL.  
 FT TRANSMEM 10799 10819 POTENTIAL.  
 FT TRANSMEM 10847 10867 POTENTIAL.  
 FT TRANSMEM 10899 10919 POTENTIAL.  
 FT TRANSMEM 10947 10967 POTENTIAL.  
 FT TRANSMEM 10999 11019 POTENTIAL.  
 FT TRANSMEM 11047 11067 POTENTIAL.  
 FT TRANSMEM 11099 11119 POTENTIAL.  
 FT TRANSMEM 11147 11167 POTENTIAL.  
 FT TRANSMEM 11199 11219 POTENTIAL.  
 FT TRANSMEM 11247 11267 POTENTIAL.  
 FT TRANSMEM 11299 11319 POTENTIAL.  
 FT TRANSMEM 11347 11367 POTENTIAL.  
 FT TRANSMEM 11399 11419 POTENTIAL.  
 FT TRANSMEM 11447 11467 POTENTIAL.  
 FT TRANSMEM 11499 11519 POTENTIAL.  
 FT TRANSMEM 11547 11567 POTENTIAL.  
 FT TRANSMEM 11599 11619 POTENTIAL.  
 FT TRANSMEM 11647 11667 POTENTIAL.  
 FT TRANSMEM 11699 11719 POTENTIAL.  
 FT TRANSMEM 11747 11767 POTENTIAL.  
 FT TRANSMEM 11799 11819 POTENTIAL.  
 FT TRANSMEM 11847 11867 POTENTIAL.  
 FT TRANSMEM 11899 11919 POTENTIAL.  
 FT TRANSMEM 11947 11967 POTENTIAL.  
 FT TRANSMEM 11999 12019 POTENTIAL.  
 FT TRANSMEM 12047 12067 POTENTIAL.  
 FT TRANSMEM 12099 12119 POTENTIAL.  
 FT TRANSMEM 12147 12167 POTENTIAL.  
 FT TRANSMEM 12199 12219 POTENTIAL.  
 FT TRANSMEM 12247 12267 POTENTIAL.  
 FT TRANSMEM 12299 12319 POTENTIAL.  
 FT TRANSMEM 12347 12367 POTENTIAL.  
 FT TRANSMEM 12399 12419 POTENTIAL.  
 FT TRANSMEM 12447 12467 POTENTIAL.  
 FT TRANSMEM 12499 12519 POTENTIAL.  
 FT TRANSMEM 12547 12567 POTENTIAL.  
 FT TRANSMEM 12599 12619 POTENTIAL.  
 FT TRANSMEM 12647 12667 POTENTIAL.  
 FT TRANSMEM 12699 12719 POTENTIAL.  
 FT TRANSMEM 12747 12767 POTENTIAL.  
 FT TRANSMEM 12799 12819 POTENTIAL.  
 FT TRANSMEM 12847 12867 POTENTIAL.  
 FT TRANSMEM 12899 12919 POTENTIAL.  
 FT TRANSMEM 12947 12967 POTENTIAL.  
 FT TRANSMEM 12999 13019 POTENTIAL.  
 FT TRANSMEM 13047 13067 POTENTIAL.  
 FT TRANSMEM 13099 13119 POTENTIAL.  
 FT TRANSMEM 13147 13167 POTENTIAL.  
 FT TRANSMEM 13199 13219 POTENTIAL.  
 FT TRANSMEM 13247 13267 POTENTIAL.  
 FT TRANSMEM 13299 13319 POTENTIAL.  
 FT TRANSMEM 13347 13367 POTENTIAL.  
 FT TRANSMEM 13399 13419 POTENTIAL.  
 FT TRANSMEM 13447 13467 POTENTIAL.  
 FT TRANSMEM 13499 13519 POTENTIAL.  
 FT TRANSMEM 13547 13567 POTENTIAL.  
 FT TRANSMEM 13599 13619 POTENTIAL.  
 FT TRANSMEM 13647 13667 POTENTIAL.  
 FT TRANSMEM 13699 13719 POTENTIAL.  
 FT TRANSMEM 13747 13767 POTENTIAL.  
 FT TRANSMEM 13799 13819 POTENTIAL.  
 FT TRANSMEM 13847 13867 POTENTIAL.  
 FT TRANSMEM 13899 13919 POTENTIAL.  
 FT TRANSMEM 13947 13967 POTENTIAL.  
 FT TRANSMEM 13999 14019 POTENTIAL.  
 FT TRANSMEM 14047 14067 POTENTIAL.  
 FT TRANSMEM 14099 14119 POTENTIAL.  
 FT TRANSMEM 14147 14167 POTENTIAL.  
 FT TRANSMEM 14199 14219 POTENTIAL.  
 FT TRANSMEM 14247 14267 POTENTIAL.  
 FT TRANSMEM 14299 14319 POTENTIAL.  
 FT TRANSMEM 14347 14367 POTENTIAL.  
 FT TRANSMEM 14399 14419 POTENTIAL.  
 FT TRANSMEM 14447 14467 POTENTIAL.  
 FT TRANSMEM 14499 14519 POTENTIAL.  
 FT TRANSMEM 14547 14567 POTENTIAL.  
 FT TRANSMEM 14599 14619 POTENTIAL.  
 FT TRANSMEM 14647 14667 POTENTIAL.  
 FT TRANSMEM 14699 14719 POTENTIAL.  
 FT TRANSMEM 14747 14767 POTENTIAL.  
 FT TRANSMEM 14799 14819 POTENTIAL.  
 FT TRANSMEM 14847 14867 POTENTIAL.  
 FT TRANSMEM 14899 14919 POTENTIAL.  
 FT TRANSMEM 14947 14967 POTENTIAL.  
 FT TRANSMEM 14999 15019 POTENTIAL.  
 FT TRANSMEM 15047 15067 POTENTIAL.  
 FT TRANSMEM 15099 15119 POTENTIAL.  
 FT TRANSMEM 15147 15167 POTENTIAL.  
 FT TRANSMEM 15199 15219 POTENTIAL.  
 FT TRANSMEM 15247 15267 POTENTIAL.  
 FT TRANSMEM 15299 15319 POTENTIAL.  
 FT TRANSMEM 15347 15367 POTENTIAL.  
 FT TRANSMEM 15399 15419 POTENTIAL.  
 FT TRANSMEM 15447 15467 POTENTIAL.  
 FT TRANSMEM 15499 15519 POTENTIAL.  
 FT TRANSMEM 15547 15567 POTENTIAL.  
 FT TRANSMEM 15599 15619 POTENTIAL.  
 FT TRANSMEM 15647 15667 POTENTIAL.  
 FT TRANSMEM 15699 15719 POTENTIAL.  
 FT TRANSMEM 15747 15767 POTENTIAL.  
 FT TRANSMEM 15799 15819 POTENTIAL.  
 FT TRANSMEM 15847 15867 POTENTIAL.  
 FT TRANSMEM 15899 15919 POTENTIAL.  
 FT TRANSMEM 15947 15967 POTENTIAL.  
 FT TRANSMEM 15999 16019 POTENTIAL.  
 FT TRANSMEM 16047 16067 POTENTIAL.  
 FT TRANSMEM 16099 16119 POTENTIAL.  
 FT TRANSMEM 16147 16167 POTENTIAL.  
 FT TRANSMEM 16199 16219 POTENTIAL.  
 FT TRANSMEM 16247 16267 POTENTIAL.  
 FT TRANSMEM 16299 16319 POTENTIAL.  
 FT TRANSMEM 16347 16367 POTENTIAL.  
 FT TRANSMEM 16399 16419 POTENTIAL.  
 FT TRANSMEM 16447 16467 POTENTIAL.  
 FT TRANSMEM 16499 16519 POTENTIAL.  
 FT TRANSMEM 16547 16567 POTENTIAL.  
 FT TRANSMEM 16599 16619 POTENTIAL.  
 FT TRANSMEM 16647 16667 POTENTIAL.  
 FT TRANSMEM 16699 16719 POTENTIAL.  
 FT TRANSMEM 16747 16767 POTENTIAL.  
 FT TRANSMEM 16799 16819 POTENTIAL.  
 FT TRANSMEM 16847 16867 POTENTIAL.  
 FT TRANSMEM 16899 16919 POTENTIAL.  
 FT TRANSMEM 16947 16967 POTENTIAL.  
 FT TRANSMEM 16999 17019 POTENTIAL.  
 FT TRANSMEM 17047 17067 POTENTIAL.  
 FT TRANSMEM 17099 17119 POTENTIAL.  
 FT TRANSMEM 17147 17167 POTENTIAL.  
 FT TRANSMEM 17199 17219 POTENTIAL.  
 FT TRANSMEM 17247 17267 POTENTIAL.  
 FT TRANSMEM 17299 17319 POTENTIAL.  
 FT TRANSMEM 17347 17367 POTENTIAL.  
 FT TRANSMEM 17399 17419 POTENTIAL.  
 FT TRANSMEM 17447 17467 POTENTIAL.  
 FT TRANSMEM 17499 17519 POTENTIAL.  
 FT TRANSMEM 17547 17567 POTENTIAL.  
 FT TRANSMEM 17599 17619 POTENTIAL.  
 FT TRANSMEM 17647 17667 POTENTIAL.  
 FT TRANSMEM 17699 17719 POTENTIAL.  
 FT TRANSMEM 17747 17767 POTENTIAL.  
 FT TRANSMEM 17799 17819 POTENTIAL.  
 FT TRANSMEM 17847 17867 POTENTIAL.  
 FT TRANSMEM 17899 17919 POTENTIAL.  
 FT TRANSMEM 17947 17967 POTENTIAL.  
 FT TRANSMEM 17999 18019 POTENTIAL.  
 FT TRANSMEM 18047 18067 POTENTIAL.  
 FT TRANSMEM 18099 18119 POTENTIAL.  
 FT TRANSMEM 18147 18167 POTENTIAL.  
 FT TRANSMEM 18199 18219 POTENTIAL.  
 FT TRANSMEM 18247 18267 POTENTIAL.  
 FT TRANSMEM 18299 18319 POTENTIAL.  
 FT TRANSMEM 18347 18367 POTENTIAL.  
 FT TRANSMEM 18399 18419 POTENTIAL.  
 FT TRANSMEM 18447 18467 POTENTIAL.  
 FT TRANSMEM 18499 18519 POTENTIAL.  
 FT TRANSMEM 18547 18567 POTENTIAL.  
 FT TRANSMEM 18599 18619 POTENTIAL.  
 FT TRANSMEM 18647 18667 POTENTIAL.  
 FT TRANSMEM 18699 18719 POTENTIAL.  
 FT TRANSMEM 18747 18767 POTENTIAL.  
 FT TRANSMEM 18799 18819 POTENTIAL.  
 FT TRANSMEM 18847 18867 POTENTIAL.  
 FT TRANSMEM 18899 18919 POTENTIAL.  
 FT TRANSMEM 18947 18967 POTENTIAL.  
 FT TRANSMEM 18999 19019 POTENTIAL.  
 FT TRANSMEM 19047 19067 POTENTIAL.  
 FT TRANSMEM 19099 19119 POTENTIAL.  
 FT TRANSMEM 19147 19167 POTENTIAL.  
 FT TRANSMEM 19199 19219 POTENTIAL.  
 FT TRANSMEM 19247 19267 POTENTIAL.  
 FT TRANSMEM 19299 19319 POTENTIAL.  
 FT TRANSMEM 19347 19367 POTENTIAL.  
 FT TRANSMEM 19399 19419 POTENTIAL.  
 FT TRANSMEM 19447 19467 POTENTIAL.  
 FT TRANSMEM 19499 19519 POTENTIAL.  
 FT TRANSMEM 19547 19567 POTENTIAL.  
 FT TRANSMEM 19599 19619 POTENTIAL.  
 FT TRANSMEM 19647 19667 POTENTIAL.  
 FT TRANSMEM 19699 19719 POTENTIAL.  
 FT TRANSMEM 19747 19767 POTENTIAL.  
 FT TRANSMEM 19799 19819 POTENTIAL.  
 FT TRANSMEM 19847 19867 POTENTIAL.  
 FT TRANSMEM 19899 19919 POTENTIAL.  
 FT TRANSMEM 19947 19967 POTENTIAL.  
 FT TRANSMEM 19999 20019 POTENTIAL.  
 FT TRANSMEM 20047 20067 POTENTIAL.  
 FT TRANSMEM 20099 20119 POTENTIAL.  
 FT TRANSMEM 20147 20167 POTENTIAL.  
 FT TRANSMEM 20199 20219 POTENTIAL.  
 FT TRANSMEM 20247 20267 POTENTIAL.  
 FT TRANSMEM 20299 20319 POTENTIAL.  
 FT TRANSMEM 20347 20367 POTENTIAL.  
 FT TRANSMEM 20399 20419 POTENTIAL.  
 FT TRANSMEM 20447 20467 POTENTIAL.  
 FT TRANSMEM 20499 20519 POTENTIAL.  
 FT TRANSMEM 20547 20567 POTENTIAL.  
 FT TRANSMEM 20599 20619 POTENTIAL.  
 FT TRANSMEM 20647 20667 POTENTIAL.  
 FT TRANSMEM 20699 20719 POTENTIAL.  
 FT TRANSMEM 20747 20767 POTENTIAL.  
 FT TRANSMEM 20799 20819 POTENTIAL.  
 FT TRANSMEM 20847 20867 POTENTIAL.  
 FT TRANSMEM 20899 20919 POTENTIAL.  
 FT TRANSMEM 20947 20967 POTENTIAL.  
 FT TRANSMEM 20999 21019 POTENTIAL.  
 FT TRANSMEM 21047 21067 POTENTIAL.  
 FT TRANSMEM 21099 21119 POTENTIAL.  
 FT TRANSMEM 21147 21167 POTENTIAL.  
 FT TRANSMEM 21199 21219 POTENTIAL.  
 FT TRANSMEM 21247 21267 POTENTIAL.  
 FT TRANSMEM 21299 21319 POTENTIAL.  
 FT TRANSMEM 21347 21367 POTENTIAL.  
 FT TRANSMEM 21399 21419 POTENTIAL.  
 FT TRANSMEM 21447 21467 POTENTIAL.  
 FT TRANSMEM 21499 21519 POTENTIAL.  
 FT TRANSMEM 21547 21567 POTENTIAL.  
 FT TRANSMEM 21599 21619 POTENTIAL.  
 FT TRANSMEM 21647 21667 POTENTIAL.  
 FT TRANSMEM 21699 21719 POTENTIAL.  
 FT TRANSMEM 21747 21767 POTENTIAL.  
 FT TRANSMEM 21799 21819 POTENTIAL.  
 FT TRANSMEM 21847 21867 POTENTIAL.  
 FT TRANSMEM 21899 21919 POTENTIAL.  
 FT TRANSMEM 21947 21967 POTENTIAL.  
 FT TRANSMEM 21999 22019 POTENTIAL.  
 FT TRANSMEM 22047 22067 POTENTIAL.  
 FT TRANSMEM 22099 22119 POTENTIAL.  
 FT TRANSMEM 22147 22167 POTENTIAL.  
 FT TRANSMEM 22199 22219 POTENTIAL.  
 FT TRANSMEM 22247 22267 POTENTIAL.  
 FT TRANSMEM 22299 22319 POTENTIAL.  
 FT TRANSMEM 22347 22367 POTENTIAL.  
 FT TRANSMEM 22399 22419 POTENTIAL.  
 FT TRANSMEM 22447 22467 POTENTIAL.  
 FT TRANSMEM 22499 22519 POTENTIAL.  
 FT TRANSMEM 22547 22567 POTENTIAL.  
 FT TRANSMEM 22599 22619 POTENTIAL.  
 FT TRANSMEM 22647 22667 POTENTIAL.  
 FT TRANSMEM 22699 22719 POTENTIAL.  
 FT TRANSMEM 22747 22767 POTENTIAL.  
 FT TRANSMEM 22799 22819 POTENTIAL.  
 FT TRANSMEM 22847 22867 POTENTIAL.  
 FT TRANSMEM 22899 22919 POTENTIAL.  
 FT TRANSMEM 22947 22967 POTENTIAL.  
 FT TRANSMEM 22999 23019 POTENTIAL.  
 FT TRANSMEM 23047 23067 POTENTIAL.  
 FT TRANSMEM 23099 23119 POTENTIAL.  
 FT TRANSMEM 23147 23167 POTENTIAL.  
 FT TRANSMEM 23199 23219 POTENTIAL.  
 FT TRANSMEM 23247 23267 POTENTIAL.  
 FT TRANSMEM 23299 23319 POTENTIAL.  
 FT TRANSMEM 23347 23367 POTENTIAL.  
 FT TRANSMEM 23399 23419 POTENTIAL.  
 FT TRANSMEM 23447 23467 POTENTIAL.  
 FT TRANSMEM 23499 23519 POTENTIAL.  
 FT TRANSMEM 23547 23567 POTENTIAL.  
 FT TRANSMEM 23599 23619 POTENTIAL.  
 FT TRANSMEM 23647 23667 POTENTIAL.  
 FT TRANSMEM 23699 23719 POTENTIAL.  
 FT TRANSMEM 23747 23767 POTENTIAL.  
 FT TRANSMEM 23799 23819 POTENTIAL.  
 FT TRANSMEM 23847 23867 POTENTIAL.  
 FT TRANSMEM 23899 23919 POTENTIAL.  
 FT TRANSMEM 23947 23967 POTENTIAL.  
 FT TRANSMEM 23999 24019 POTENTIAL.  
 FT TRANSMEM 24047 24067 POTENTIAL.  
 FT TRANSMEM 24099 24119 POTENTIAL.  
 FT TRANSMEM 24147 24167 POTENTIAL.  
 FT TRANSMEM 24199 24219 POTENTIAL.  
 FT TRANSMEM 24247 24267 POTENTIAL.  
 FT TRANSMEM 24299 24319 POTENTIAL.  
 FT TRANSMEM 24347 24367 POTENTIAL.  
 FT TRANSMEM 24399 24419 POTENTIAL.  
 FT TRANSMEM 24447 24467 POTENTIAL.  
 FT TRANSMEM 24499 24519 POTENTIAL.  
 FT TRANSMEM 24547 24567 POTENTIAL.  
 FT TRANSMEM 24599 24619 POTENTIAL.  
 FT TRANSMEM 24647 24667 POTENTIAL.  
 FT TRANSMEM 24699 24719 POTENTIAL

UL06 HSV6Z STANDARD; PRT; 662 AA.  
ID UL06 HSV6Z  
AC P52454;  
DT 01-OCT-1996 (Rel. 34, Created)  
DT 01-OCT-1996 (Rel. 34, Last sequence update)  
DT 10-OCT-2003 (Rel. 42, Last annotation update)  
DE Virion protein U76.  
UN 776 OR CB3L.  
OS Human herpesvirus (type 6 / strain Z29) (HHV6).  
OC Betaherpesviridae; Herpesviridae;  
OC Betaherpesvirinae; Roseolovirus.  
OX NCBI\_TaxID=36351;  
RN  
RP SEQUENCE FROM N.A.  
RX MEDLINE=96195263; PubMed=8634027;  
RA Lindquester G.J., Inoue N., Allen R.D., Castelli J.W.,  
RA Stamey F.R., Dambaugh T.R., O'Brian J.J., Danovitch R.M.,  
RA Frenkel N., Pellett P.B.;  
RT "Restriction endonuclease mapping and molecular cloning of the human  
herpesvirus 6 variant B strain Z29 genome.";  
RL Arch. Virol. 141:367-379 (1996).  
RN  
RP SEQUENCE FROM N.A.  
RX MEDLINE=99412318; PubMed=10482553;  
RA Dominguez G., Dambaugh T.R., Stamey F.R., Dewhurst S., Inoue N.,  
RA Pellett P.B.;  
RT "Human herpesvirus 6B genome sequence: coding content and comparison  
with human herpesvirus 6A.";  
RL J. Virol. 73:8040-8052 (1999).  
CC -1- FUNCTION: PRESUMED VIRION PROTEIN; POSSIBLE ROLE IN DNA  
PACKAGING.  
CC  
CC -1- SIMILARITY: BELONGS TO FAMILY THAT GROUPS TOGETHER HSV-1 UL6,  
EHV-1 56, EBV BBRF1, HCMV UL104, AND VZV 54.  
CC  
CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
between the Swiss Institute of Bioinformatics and the EMBL outstation -  
the European Bioinformatics Institute. There are no restrictions on its  
use by non-profit institutions as long as its content is in no way  
modified and this statement is not removed. Usage by and for commercial  
entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
CC  
DR EMBL; AF157706; AAB06359.1; --  
DR PIR; T44221; T44221.  
DR InterPro; IPR002860; Herpes\_UL6.  
DR Pfam; PF01763; Herpes\_UL6; 1.  
DR ProDom; PD003210; Herpes\_UL6; 1.  
SQ SEQUENCE 662 AA; 77199 MW; 5126392A37C67B90 CRC64;  
Query Match 6.8%; Score 81.5; DB 1; Length 662;  
Best Local Similarity 22.2%; Pred. No. 10;  
Matches 34; Conservative 29; Mismatches 55; Indels 35; Gaps 5;  
QY 105 VFSRVNNVY-----LEAPLVGIBGLKSGTYNL-----LFCGSGGIPVGHLY 149  
DB 213 VSSKLHDIYRHQNIIVTPIILGLSSVILIDPHNIKIPDRNSEQISCFENKKAIAFFTY 272  
QY 150 STHAAL-----AALRGHCLSSDKWVCYLTKXAI--VNASEMDIQWPLS 193  
DB 273 STYVIRNRIMLTTPLAHLSPRLKKNLSLRHQKMCCLINTFPKIVLAKTVTN----- 328  
QY 194 EKTAELKEKIVLTHNRKLSIMKILSEVTPDSK 226  
DB 329 KKTMDLKEKEKNSDAKSKLIFLNLNDSKSK 361  
RESULT 12  
CALC MOUSE STANDARD; PRT; 3119 AA.  
ID CALC MOUSE  
AC Q60847; P70322;  
DT 15-JUL-1998 (Rel. 36, Created)  
DT 28-FEB-2003 (Rel. 41, Last sequence update)  
DT 10-OCT-2003 (Rel. 42, Last annotation update)

DE Collagen alpha 1(XII) chain precursor.  
GN COL12A1.  
OS Mus musculus (Mouse).  
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.  
OX NCBI\_TaxID=10090;  
RN  
RP SEQUENCE FROM N.A., AND ALTERNATIVE SPLICING (ISOFORMS XIIA-1 AND  
XIIA-2).  
RP STRAIN=Swiss Webster, and C57BL/6J; TISSUE=Skin;  
RX MEDLINE=96170761; PubMed=8601036;  
RA Boehme K., Li Y., Oh P.S., Olsen B.R.;  
RT "Primary structure of the long and short splice variants of mouse  
collagen XII and their tissue-specific expression during embryonic  
development.";  
RL dev. Dyn. 204:432-445 (1995).  
RN  
RP PARTIAL SEQUENCE FROM N.A., AND ALTERNATIVE SPLICING (ISOFORMS XIIA-2  
AND XIIA-1).  
RP STRAIN=C57BL/6J; TISSUE=Skin fibroblast;  
RX MEDLINE=99348349; PubMed=10419532;  
RA Kania A.M., Reichenberger E., Baur S.T., Karimbux N.Y., Taylor R.W.,  
RA Olsen B.R., Nishimura I.;  
RT "Structural variation of type XII collagen at its carboxyl-terminal  
NC1 domain generated by tissue-specific alternative splicing.";  
RL J. Biol. Chem. 274:22053-22059 (1999).  
CC -1- FUNCTION: Type XII collagen interacts with type I collagen-  
containing fibrils, the COL1 domain could be associated with the  
surface of the fibrils, and the COL2 and NC3 domains may be  
localized in the pericellular matrix (By similarity).  
CC -1- SUBUNIT: Trimer of identical chains each containing 190 kDa of  
nontriple-helical sequences (By similarity).  
CC -1- ALTERNATIVE PRODUCTS:  
Event=Alternative splicing; Named isoforms=4;  
Comment=The final tissue form of collagen XII may contain  
homotrimers or any combination of the various isoforms;  
Name=XIIA-1;  
Isoid=Q60847-1; Sequence=Displayed;  
Name=XIIA-2; Synonyms=ER#K;  
Isoid=Q60847-2; Sequence=VSP\_001151, VSP\_001152;  
Name=XIIB-1;  
Isoid=Q60847-3; Sequence=VSP\_001150;  
Name=XIIB-2;  
Isoid=Q60847-4; Sequence=VSP\_001150, VSP\_001151, VSP\_001152;  
CC -1- TISSUE SPECIFICITY: Highest expression in tendons, perichondrium,  
skin, cornea, sclera, blood vessels, and peritoneum.  
CC -1- DEVELOPMENTAL STAGE: The long NC3 XIIA isoforms are predominant at  
early stages (ED7 and 11); at later stages of development (ED15  
and 17) the short NC3 XIIB forms become the major forms. As the  
short NC3 forms become the major product, the long splice variant  
continues to be expressed in several tissues, even after birth.  
CC The long NC1 isoforms, XIIA-1 and XIIB-1, peak in 15-day old  
embryos and decrease in 17-day old ones. The expression of the  
short NC1 form XIIB-2 remains constant throughout late stages of  
embryonic development (ED15 and ED17).  
CC -1- PTM: The triple-helical tail is stabilized by disulfide bonds at  
each end (By similarity).  
CC -1- PTM: Prolines at the third position of the tripeptide repeating  
unit (G-X-Y) are hydroxylated in some or all of the chains (By  
similarity).  
CC -1- PTM: O-glycosylation of isoform XIIA-2; glycosaminoglycan of  
chondroitin-sulfate type (By similarity).  
CC -1- SIMILARITY: BELONGS TO THE FIBRIL-ASSOCIATED COLLAGENS WITH  
INTERRUPTED HELICES (FACIT) FAMILY.  
CC -1- SIMILARITY: Contains 1 TSP N-terminal (TSPN) domain.  
CC -1- SIMILARITY: Contains 4 WFPA domain.  
CC -1- SIMILARITY: Contains 18 fibronectin type III domains.  
CC  
CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
between the Swiss Institute of Bioinformatics and the EMBL outstation -  
the European Bioinformatics Institute. There are no restrictions on its  
use by non-profit institutions as long as its content is in no way  
modified and this statement is not removed. Usage by and for commercial

FT	CARBOHYD	704	704	N-LINKED (GLCNAC. .) (POTENTIAL).
FT	CARBOHYD	802	802	O-LINKED (XYL. .) (CHONDROITIN SULFATE)
FT	CARBOHYD	893	893	(POTENTIAL).
FT	CARBOHYD			O-LINKED (XYL. .) (CHONDROITIN SULFATE)
FT	CARBOHYD	985	985	(POTENTIAL).
FT	CARBOHYD			O-LINKED (XYL. .) (CHONDROITIN SULFATE)
FT	CARBOHYD	1769	1769	(POTENTIAL).
FT	CARBOHYD	2212	2212	N-LINKED (GLCNAC. .) (POTENTIAL).
FT	CARBOHYD	2533	2533	N-LINKED (GLCNAC. .) (POTENTIAL).
FT	CARBOHYD	2684	2684	N-LINKED (GLCNAC. .) (POTENTIAL).
FT	VARSPLIC	25	1190	Missing (in isoform XIIB-1 and isoform XIIB-2).
FT				/FTID=VSP_001150.
FT	VARSPLIC	3062	3064	Epy -> GSG (in isoform XIIA-2 and isoform XIIB-2).
FT				/FTID=VSP_001151.
FT	VARSPLIC	3065	3119	Missing (in isoform XIIA-2 and isoform XIIB-2).
FT				/FTID=VSP_001152.
SQ	SEQUENCE	3119 AA;	340239 MM;	9B1F959C86AB3251 CRC64;
Query Match		6.8%;	Score 81.5%;	DB 1; Length 3119;
Best Local Similarity		21.6%;	Pred No. 74;	Mismatches 43; Conservative 32; Mismatches 85; Indels 39; Gaps 8;
Dy	34	TTSMEHDIOYVKGSSPLPGAGIAGEEPAAPQLPSWLOPERCANVQCACQCHVLADSVHL	93	
Dy	2282	TENLEBPGPVPIKEQTIVKTEAPTPEPTPPSP--PTTIPPARDV-----CKAKRKDIYFL	2333	
Dy	94	AMDLSRSGAVVESVTNNVLEAFVINGIESLGSTYNLLFGSCGIPVGFIHLYSTHA	153	
Dy	2334	T-DASWSIGDDNFKNVR-----FIRNTGAFFDE-----VNPAIQVSFWGYSD-	2377	
Dy	154	ALALRGHCLESDDMKCYLLTKTRIVASEDDID---NVPISKIEMIEKYVLTHNR	209	
Dy	2378	----VKSEPKLNT-----YNDRAIALGALONIRYRGGNRTTGKALTFPIEKVLTWESG	2426	
Dy	210	LKSLMKILISEVTTPDOSKPE	228	
Dy		::::: ::: :::		
Dy	2427	MKNVRLGVVTDGRSODE	2445	
RESULT 13				
YN06_SYNEL				
ID_YN06_SYNEL	STANDARD;	PRT;	371 AA.	
AC_Q8DSLO;				
DT	10-OCT-2003 (Rel. 42, Created)			
DT	10-OCT-2003 (Rel. 42, Last sequence update)			
DE	Hypothetical UPF0284 protein t112306.			
OS	Synechococcus elongatus (thermosynechococcus elongatus).			
OC	Bacteria; Cyanobacteria; Chroococcales; Synechococcus.			
OX	NCBI_TaxID=32046;			
RN	[1]			
RP_SEQUENCE FROM N.A.				
RC_STRAIN=BP-1;				
RX_MEDLINE=2225144;	PubMed=12240834;			
RA_Nakamura Y., Kaneko T., Sato S., Ikeuchi M., Katoh H., Sasamoto S.,				
RA_Matsumoto A., Iriiguchi M., Kawashima K., Kimura T., Kishida Y.,				
RA_Kiyokawa C., Kohara M., Matsumoto M., Matsuno A., Nakazaki N.,				
RA_Shimo S., Sugimoto M., Takeuchi C., Yamada M., Tabata S.;				
RT"Complete genome structure of the thermophilic cyanobacterium				
Thermosynechococcus elongatus BP-1."				
DNA Res. 9:123-130(2002).				
-1- SIMILARITY: Belongs to the UPF0284 family.				
CC -----				
CC This SWISS-PROT entry is copyright. It is produced through a collaboration				
CC between the Swiss Institute of Bioinformatics and the EMBL outstation				
CC at European Bioinformatics Institute. There are no restrictions on its				
CC use by non-profit institutions as long as its content is in no way				



CC modified and this statement is not removed. Usage by and for commercial  
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
CC -----  
DR EMBL; Z22177; CAA80145.1; -  
DR PIR; S40764; S40764.  
DR Wormped; ZK512.3; CE00409.  
DR InterPro; IPR002111; Cat\_channel TprL.  
DR SEQUENCE 1418 AA; 159237 MW; 4FC83B9E7ADE7788 CRC64;  
SQ  
Query Match 6.8%; Score 81; DB 1; Length 1418;  
Best Local Similarity 28.3%; Pred. No. 30;  
Matches 32; Conservative 15; Mismatches 36; Indels 30; Gaps 5;  
QY 118 PFLVGI-EGSLKSGSTYNLFCGSCGIPVG-----HYSTHAALALRGHFCLSSD 167  
Db 230 PFPVAFPAAGASKESLIELFVEHGIPIVILQDSCELCALIHSHLLETNS---FDND 285  
QY 168 KMV-----CYLKTKAIVNASEMDIQNVPL--SEKIAELKEKIV 204  
Db 286 KFI\$WLR\$QULYPLGLADCYTLITKLIVSNSGDVQLIEFIDSSQLSELSSVW 338

Search completed: July 20, 2004, 11:16:53  
Job time : 15 secs